



Bringing Big Data to Agriculture, and Agriculture to Big Data



CGIAR

Science for a food-secure future

VISIT...

LANZAROTE
Caliente.COM

CGIAR Big Data Coordination Platform

**Leveraging CGIAR data:
Bringing big data to agriculture,
and agriculture to big data**

July 2016

This proposal was developed from consultations with CGIAR CRPs and Centers and with representatives of 42 external partners. CIAT and IFPRI will circulate it more widely seeking further inputs from institutions and potential partners in the public and private sector with interest in big data and its potential contribution to development of agriculture to ensure global food security.

Citation: CIAT and IFPRI (2016). CGIAR Big data coordination platform. Proposal to the CGIAR Fund Council, 31 March, 2016. International Center for Tropical Agriculture (CIAT), Cali, Colombia and International Food Policy Research Institute, Washington DC, United States of America.

CGIAR Big data coordination platform

Leveraging CGIAR data: Bringing big data to agriculture, and agriculture to big data

CONTENTS

1.0 Platform Narrative	1
1.0.1 Rationale and scope.....	1
1.02 Platform Structure and Modules	5
1.0.3 Impact pathway and contribution to theory of change of the CGIAR and its programs.....	7
1.0.4 Capacity development	9
1.0.5 Platform leadership, management and governance	10
1.1 Platform Budget Narrative	12
1.1.1 General Information	12
1.1.2 Summary.....	12
1.1.3 CRP Funding Plan	13
1.1.4 CRP Management and Support Costs.....	14
1.1.5 CRP Financial management principles.....	15
1.1.6 Budgeted costs for key activities:	15
2 Business Case	17
2.1 Expertise and track record.....	17
2.2 Access to Materials, Sites and Services	21
2.3 Interactions between the Platform and Users	21
2.4 Intellectual Asset Management.....	22
2.5 Open Access Management	22
2.6 Communication Strategy	23
2.7 Risk Management	23
3. MODULES	24
3.1 MODULE 1: ORGANIZE.....	24
3.1.1 Modules Narrative.....	24
3.1.2 Module Budget Narrative	34
3.1.3 Module Uplift Budget	40
MODULE 2: CONVENE.....	45
3.2.1 Module Narrative	45
3.2.2 Module Budget Narrative	51

3.2.3 Module Uplift Budget	54
3.3 MODULE 3: INSPIRE	56
3.3.1 Module Narrative	56
3.3.2 Module Budget Narrative	62
3.3.3 Module Uplift Budget	65
4. ANNEXES	66
4.1 ANNEX 1 EXPERTISE AND TRACK RECORD - CVS	67
4.2 ANNEX 2 BUSINESS CASE: INTERACTIONS BETWEEN THE PLATFORM AND USERS	84
4.3 ANNEX 3. OPEN ACCESS (OA) AND OPEN DATA (OD) MANAGEMENT	85
Planning for and Implementing OA/OD in Accordance with the CGIAR OADM Policy and FAIR Principles, including Critical Issues and Anticipated Challenges	85
Technical Considerations and Operations	86
Coordination and Decision Making	87
Narrative for Required Resources	88
4.4 ANNEX 4. INTELLECTUAL ASSET MANAGEMENT	89
Overview of intellectual assets management in the Platform	89
Planning and tracking concerning intellectual assets and their associated rights	89
Indicative budget breakdown by year relating to intellectual assets management.	91
4.6 ANNEX 5 - LIST OF REFERENCES	92
4.7 ANNEX 6 - PLATFORM LEADERSHIP, MANAGEMENT AND GOVERNANCE (DETAILED)	94
4.8 ANNEX 7. CONSULTATIONS: NEEDS ASSESSMENTS AND GAPS	97
AMBITION	97
INSPIRE	97
GOVERNANCE	98
EXPERTISE AND PARTNERSHIPS	99
TOOLS	99
SUMMARY	101
4.8 ANNEX 8. Ethical considerations in relation to the requirement for IRB approval	102
4.9 ANNEX 9. Curation, quality review and certification of data	103

1.0 PLATFORM NARRATIVE

1.0.1 Rationale and scope

Overview and Development Context

Data has become a valuable global commodity. But it is much more than simply information: in expert hands, it is intelligence. Already it is helping accelerate the development of robust responses to some of the most pressing challenges of our time: climate change/variability, food insecurity and malnutrition, and environmental degradation. It is transforming the world of genomics and crop breeding and revolutionizing disciplines from climate modelling to agronomy. It is helping refine policies and improve lives. The smart/wise/effective use of data will be one of the most important tools for achieving the United Nations' Sustainable Development Goals.

The technical definition of big data varies across disciplines, but is generally characterized as having high volume, velocity, variety, and variability (Laney 2001). For CGIAR, we loosely define big data as harmonized, interoperable, and contextually integrated datasets and publications from multiple disciplines relevant for CGIAR's research and development goals (CGIAR Consortium 2015a). The applications of insights gained from the study of such integrated datasets are greatly advancing knowledge and impact in several fields, including the -omics, biomedical, and ecological domains (Kitchin 2014; Madin et al. 2008; Jin et al. 2015).

The concept of data being a driving force to enhance agricultural competitiveness is clearly on the radar screen of agribusiness private sector (Bryceson 2006, Hardy 2013, Gilpin 2014). For years, this has manifested itself as precision farming (Zhang et al. 2002), and been limited in scope to optimal management of within plot productivity variability. But advances in recent years have significantly opened the scope for data driven agricultural development (Jin et al. 2015). The high profile US\$1bn acquisition of Climate Corporation by Monsanto (Monsanto 2013) demonstrated that information can be as valued a product as a novel seed technology, and has spurred significant investment and development of information-based (i.e., data and decision support-based) agricultural services. Data driven approaches to farming in urban areas (e.g. vertical farming) are showing significant growth, and science fiction is coming reality with pilots of [robot-farming](#).

The opportunities that the data revolution offer have the potential to change the role, reach and modus operandi of research and development organizations such as CGIAR (Seck et al. 2013). It represents an unprecedented opportunity to find new ways of reducing hunger and poverty, by encompassing data-driven solutions to ongoing research for development efforts and in so doing, bringing transformational changes by introducing positive disruptive technologies that radically change impact pathways (Seck et al. 2013). But the data revolution also has risks: unequal access to and use of information could widen social inequity, and exacerbate yield gaps in agriculture (Warschauer 2004). The ability to generate, analyze and use data is weighted towards resource-rich farmers and their supporting institutions in developed countries (Hardy 2013; Gilpin 2014). Access to smart-phones, ICT technologies, and mobile data and private sector providers is significantly weaker in developing versus developed countries. While African yields have stagnated over the past 2 decades, productivity in developed countries continue to increase thanks to access to agricultural technology, precision agriculture and strong institutions. Crop yields in many regions approach their practical maxima (Fischer et al. 2014). The need to feed the projected world population of almost 10 billion in 2050 will have to be met by yields in rain-fed agriculture, especially in Africa bringing them

closer to their practical maxima (Rost et al. 2009). At present, they achieve only a low fraction of that level (Fischer et al. 2014).

In developing countries, there are already a number of success stories. In Kenya, [identified as a hub for novel ICT-based innovations](#), farmers can gain access to advisory information through services such as [iCow](#). These are transforming information access to resource-poor farmers and are changing age-old concepts of agricultural extension. Similarly in Mexico the technology transformation program [MASAGRO](#) run by the Mexican government with CIMMYT, has made satellite imagery available in a [tool](#) for calculating nitrogen requirements of farmer fields.

The capacity to innovate around problems in developing countries is clearly there – mobile money in Kenya has revolutionized banking in urban and rural settings alike (Demombynes and Thegeya 2012), and has shamed developed countries with simple, safe and effective banking that is accessible to all. This Platform dreams about generating equally impactful big data and ICT innovations that revolutionizes farming in developing countries. CGIAR is uniquely positioned to be a thought leader on the use of big data and information technology in agriculture to drive equitable rural development. In so doing, CGIAR can help ensure that the data revolution is democratic, and reaches the marginalized rural poor.

There is a potential to enhance agricultural innovation and impact using big data and contributing to impacts outlined in CGIAR's [Strategic Results Framework](#) (SRF). The Big Data Coordination Platform (the Platform hereafter) will provide a strategic collaboration encompassing all CGIAR Centers, Programs, and key partners across public and private sectors. It will provide useful backend services and opportunities for researchers to analyze and visualize on a massive scale agricultural data, with which it is otherwise difficult to generate rapid, actionable insights.

Objectives and Targets

This proposal for a Big Data and ICT Platform therefore focuses on enhancing CGIAR and partner capacity to deliver big data management, analytics and ICT-focused solutions to CGIAR target geographies and communities. The ultimate goal of the platform is **to harness the capabilities of Big Data to accelerate and enhance the impact of international agricultural research**. It will support CGIAR's mission by creating an enabling environment where data are expertly managed and used effectively to strengthen delivery on CGIAR SRF's System Level Outcome (SLO) targets. Critical gaps were identified during the extensive scoping consultations with CGIAR researchers and partners (provided in Annex 8).

The Platform will achieve this through ambitious partnerships with initiatives and organizations outside CGIAR, both upstream and downstream, public and private. It will focus on promoting CGIAR-wide collaboration across CRPs and Centers, in addition to developing new partnership models with big data leaders at the global level. As a result, CGIAR and partner capacity will be enhanced, external partnerships will be leveraged, and an institutional culture of collaborative data management and analytics will be established. Important international public goods such as new global and regional datasets will be developed, alongside new methods that support CGIAR to use the data revolution as an additional means of delivering on SLOs.

The Platform has the following three objectives:

1. **Support and improve data generation, access, and management in CGIAR:** For CGIAR to embrace the power of big data analytics and be the leader in generating actionable data-driven insights for stakeholders, key requirements, enabling environment components, and critical gaps, which were identified during the scoping consultations. Closely collaborating

with partners identified as the champions in bringing big data to agriculture, the Platform will provide support to CGIAR and partners to address the gaps, both organizational (i.e., OA/OD-compliance) and technical (e.g., providing useful datasets, tools, and services), and organize capacity building activities to sustain the efforts across the consortium.

- Collaborate and convene around big data and agricultural development: CGIAR needs ambitious external partnerships to deliver the potential of big data to smallholder agriculture. Likewise, CGIAR is an attractive boundary partner for many private and public big data partners to engage in the context of agriculture in the developing world. This objective will set up system-level partnerships that Centers and CRPs can tap into and use to stimulate greater use of data analytics in CGIAR mission-critical research. Amongst other approaches, the Platform will provide opportunities and spaces for facilitated virtual collaboration and interaction among partners and stakeholders. A Big Data Convention will be organized to bring key actors to CGIAR and CGIAR to the key actors in a network that will be documented and nurtured. The Convention will focus on the generation of ideas and innovations. It will democratize big data opportunities, share progress amongst CRPs and Centers in promoting big data analytics. It will build capacity internally and externally on big data approaches in agriculture. Novel approaches to communications will increase exposure of CGIAR work on big data, and further engage a range of actors through novel approaches to partnerships.
- 2. **Lead by example and inspire how big data can deliver development outcomes:** Demonstrate the power of CGIAR big data analytics through “Inspire” projects that solve development challenges at the core of CGIAR SRF. These may include, but not be exclusive to, approaches that use big data analytics and ICTs to provide unprecedented multi-disciplinary data to researchers, deliver novel information to farmers, monitor the state of agriculture and food security in real time and inform critical national, regional and global policies and decisions. Venture capital (<\$100k) will be provided to generate novel approaches, and some larger projects will be developed to deliver on the overall vision of the Platform: democratize big data to include smallholder farmers.

These three objectives are sequenced purposefully. As a first and necessary step, CGIAR must get its own house in order with respect to data. A significant step forward is required that values data as a product in itself with global public good potential, and uses the best available means of managing it and making it available to the public following OA/OD principles. This requires the infrastructure, tools and data culture to succeed, in both technical and managerial dimensions, and **Objective 1** focuses precisely on this. The minimum success factor for this objective is compliance with the open access and open data policy of CGIAR, ensuring donors and investors in CGIAR can be confident that data is being managed and shared effectively across all CGIAR operations. **Objective 2** enables the external partnerships needed to deliver on Objective 1, and aims to enable CGIAR to make a significant advance in its capacity to manage, use and analyze data. New partnership models will be developed with upstream and downstream partners, from public and private sectors, to deepen and widen CGIAR's capacity on big data analytics and use. Finally, **Objective 3** is designed to inspire through the operationalization of CGIAR and partners capacity to innovate around big data. It aims to tackle some of development's most complex problems with new data-driven approaches, and deliver scalable pilots of effective big data for development solutions. Established models for innovation such as Grand Challenges of the Bill and Melinda Gates Foundation (Cohen 2005) and other approaches will be used to achieve this, built on the foundation of collaboration from Objective 2.

Platform Principles

Process-oriented agile approach

Rather than a comprehensive plan of very specific activities for the Platform that describes a sequential pre-defined development over a 6 year period, this proposal focuses on establishing the processes and collaborative spaces needed to deliver these goals in phases. The approach favors agility and iterative interactions with users that will bring enough flexibility to adopt emerging technologies to fulfill rising needs.

Networked approach through partnership

As problems in society have become more complex and the nature of solutions more independent, Extension systems have had greater and greater pressure to work across traditional expertise, as well as geographic and institutional boundaries to leverage maximum impact from resources. As this pressure grows, networks and communities will need to emerge as the basis for solving problems rather than single institutions. These communities of practice can leverage technology and new data resources to create broader and deeper impact in programming. Enabling communities of practice to thrive through connectivity, sharing of methodologies, knowledge and by telling their impact stories is the long term goal.

The Platform is therefore to work seamlessly across CGIAR CRPs and Centers through a networked approach.

Iterative DATA needs assessment and TECHNOLOGY landscape analysis

Big data technologies and initiatives are highly dynamic and the landscape is changing at rapid rates. A thorough data landscape analysis will be conducted regularly to understand what Open Data initiatives and Big Data based ICT4D initiatives already exist, such as [Open Data for Africa](#) from AfDB, and analyze if they are used, and their potential needs and current data gaps. The analysis will also define potential business cases where new data-driven insights could help raise investments in agriculture and guide business decisions, from product innovation to market access and infrastructure development. This assessment will be iteratively implemented on a regular basis (i.e., annually or biennially), to keep the Platform's relevance for the newly emerging agricultural research and development topics and big data technologies.

Variety is also the most important dimension of big data in agriculture, and this applies beyond the well-structured datasets. A key aspect to achieve success in the application of big data analytics is the ability to identify, combine, and efficiently manage multiple data – both from the conventional structured databases and unconventional unstructured data streams. However, it is not straightforward when identifying which data are potentially useful for what types of research questions, analyses and appropriate large data sets. For this, the Platform will work closely with partners in public (e.g., Ministry of Agriculture) and private (e.g., key players in priority commodity's value-chain) sectors in CGIAR site integration countries in Africa, Asia, and Latin America, and regularly assess primary user needs through a multi-partner, multi-datastream, multi-country project in each region.

Extensive consultations were held in the development of this proposal with CGIAR partners, CRPs and Centers to help identify the focus and contents of this Platform. Details of these consultations are available in Annex 8. Constant consultation with users (CRPs, centers, partners and others) will be maintained throughout the platform.

1.02 Platform Structure and Modules

Structure

CIAT and IFPRI will provide overall co-leadership, but a partnership model built on a networked infrastructure will be promoted, whereby the Platform plays a facilitation approach to bringing CRPs, Centers and external partners together toward focused problem solving. The Platform is submitted on behalf of about 50 external partners, in addition to all 12 CRPs and 15 Centers. Additional partners who express interest in participating and bring additional expertise to the Platform will be welcomed.

The Platform will be facilitated through a small secretariat (3 people, US\$200k annual budget) with the primary objective of facilitating dialogue, collaboration and communication across CGIAR and with external partners. The Platform Secretariat would take a knowledge sharing approach to interacting and networking amongst the partners.

Modules

The Platform proposes three solution-oriented Modules, which follow the three objectives outlined in the previous section.

Module #1: Organize

A wide range of data and analytics opportunities are becoming available in CGIAR and its partners, yet not many of them are fully utilized for developing insights and impact due to inadequate data management. Based on the thorough assessment on the needs and gaps, the Organize Module will first aim to fully open access to CGIAR's intellectual assets, addressing technical and organizational challenges, and provide CGIAR researchers with enabling environment components to strengthen data analytical capacity and develop practical, big data-driven use-cases in a coordinated way. As a minimum, this Module will align CGIAR Centers on open access and open data, and ensure compliance with CGIAR's Open Access Policy, ratified in late 2013 by all 15 Centers.

Module #2: Convene

The Platform will implement an annual CGIAR Big Data and ICT Convention where representatives from Centers and CRPs will share information, develop joint initiatives and collaborate with key external actors in the big data space. It will establish Communities of Practice (CoPs) across Centers for defining data standards and interoperability protocols, dovetailed with the OA/OD initiatives of the Consortium Office. It will create virtual collaboration spaces, and sponsor quarterly webinars on specific topics and hold capacity-building workshops. It will connect with existing initiatives, other GIAR CoPs, Conventions on ICT4D and Big Data. The Platform will also connect CGIAR with existing initiatives, for instance, the Global Agricultural Concept Scheme (GACS), FAO's AIMS, CIARD, the Interest Group on Agricultural Data (IGAD) of the Research Data Alliance (RDA), and the Global Open Data in Agriculture and Nutrition (GODAN) initiative are of high importance and relevance.

Module #3: Inspire

In order to generate high profile, collaborative applications of big data in agriculture, small- and medium- sized "Inspire" projects will be implemented that bring to bear new partnerships on core CGIAR challenges. These will consist of 1-3 year initiatives to generate new analytical approaches, scientific outputs and high profile examples of how big data analytics can deliver agricultural development in CGIAR target geographies. These projects will be embedded within larger CRP-related initiatives to leverage effectively other CGIAR investments, and will bring new and exciting partnerships with some of the major global players in big data globally and regionally.

The Platform and Big Data in CGIAR

Together the three Modules attack the most important facets of the problem that big data present to CGIAR. The first **organizes** the data by finding out what data exist and draws them together for unified and interlinked discoverability. It assesses their status and fitness for use (quality metadata/annotations, similar spatial and temporal scales, etc.), identifies what and where gaps exist, and strengthens its analytical capacity for data-driven impact. It seeks to involve a wide range of Platform partners to produce tools that will allow the users, scientists, policy makers, or farmers to access the data in transparent, easy-to-use ways, and to support synthesizing data-driven insights. The second **convenes** the scientific resources across CGIAR with a vast range of partners to generate new collaborative opportunities and bring big data to agriculture, and likewise, agriculture to big data. The third **inspires** work on big data by funding research by CGIAR scientists with partners to innovate new ways to handle big data relevant to agriculture for the benefit of poor smallholder farmers.

The Platform's Linkage with Other CGIAR Platforms and Programs

The **Organize** Module will actively work with Center and CRP researchers and CoPs (including the data, knowledge, IT, legal, and other relevant system-wide CoPs) in the inventory and management of datasets towards "open", and supporting tools for researchers' use. The goal is to support all Centers and CRPs to not only comply but overachieve with regard to open access and open data principles and CGIAR policy on these. The **Convene** Module will bring together big data practitioners from Centers and CRPs together with partners and other Platforms in spaces that will encourage interaction and pursuit of common goals. We expect that the interactions will produce ideas that we will encourage to apply for funding under the **Inspire** Module to develop them. The **Inspire** Module will create opportunities for novel ideas to be realized through pilot projects with new partners to CGIAR under collaborative efforts embedded within CRP activities.

Collaboration is envisioned with the Genetic Gains Platform through shared infrastructure for processing of genetic data, and the Genebank Platform to the extent to which environmental and socio-economic data can add value to passport information for germplasm collections. The Platform will collaborate with CGIAR Gender research network, for example, on the adoption of standards for sex-disaggregated data collection and analysis (Doss and Kieran, 2013). Cross-program collaboration will be facilitated in gender research for data warehousing and data sharing, including a protocol for sharing in compliance with Open Access. The members of the Capacity Development CoP will also be a crucial human resource to design, develop, deliver, monitor and evaluate the capacity development efforts of the Platform.

1.0.3 Impact pathway and contribution to theory of change of the CGIAR and its programs

The Platform aims to increase the impact of agricultural development by embracing big data and ICT approaches to ***solve development problems faster, better and at greater scale***. As outlined in the SRF, this will initially be across CGIAR, but will be extensible to agriculture at large.

The Theory of Change of the Platform focuses on increasing the capacity of CGIAR and partners to embrace big data and ICT approaches. This is achieved through:

- **Enhanced collaboration** between CRPs and Centers on data analytics and ICTs;
- **Enabling unrestricted discoverability of inter-linked data sets** to understand and tackle multi-faceted challenges in data in evidence-driven ways;
- **Leveraging CGIAR expertise** with external partners to broaden and deepen experience and capacity and **position CGIAR** in the broader big data and ICT sphere to establish CGIAR as an innovative thought leader; and
- Development initiatives using proven big data innovations to drive agricultural growth in developing countries.

To enable the above, each CGIAR Center's data will be collected, stored, documented, and made available. The Platform will focus on supporting CRPs and Centers to comply with CGIAR's Open Access and Data Management (OA/OD) Policy. This will satisfy its investors that CGIAR is able to deliver data-related global public goods. A necessary first step in this theory of change is to enhance CGIAR's capacity to produce high-quality data products and provide open access to them. Readily-available data spurs their greater use. Through partnerships with public and private, upstream and downstream partners, CGIAR will become a broker of big data information, which will promote data-driven agricultural development. The Platform will achieve this by developing novel methodologies and innovative pilots, implemented within the context of CGIAR investment in CRPs. Public and private partners that are not traditional to CGIAR are key to achieving this. The partnerships will bring together institutions with complementary big data expertise ranging from research (e.g. high end data analytics) to development (provide ICT-based services to beneficiaries). Connecting CGIAR scientists to the network of big data practitioners and developers will expand the delivery capability and horizons of CGIAR research.

The expected outcomes of the Platform include:

- CGIAR researchers are more familiar with ongoing work across the system related to big data and ICTs, and greater data and knowledge sharing.
- Delivery on CGIAR's OA/OD policy that currently underutilized data are discovered, integrated, and reused to achieve development outcomes.
- Increased multi-Center, multi-CRP projects across CGIAR on big data and ICT with less redundancy in activities.
- High-profile media coverage on big data and ICT success stories from CGIAR.
- Recognition of CGIAR as a global thought leader on big data and ICTs in agriculture and development.

Although CRPs are the primary contributors to CGIAR IDOs, the Platform aims to increase the likelihood of reaching their targets. It will also generate capacity within CGIAR in the medium-long term by to reach scale with its research. The Inspire Projects will contribute to specific IDOs and sub-IDOs. Big data and ICT solutions will contribute to gender and youth IDOs. It will provide technology

that enables women and youth to access and use information that improve farming and livelihoods. Figure 1 shows how Platform outputs transfer to outcomes and relate to CGIAR IDOs.

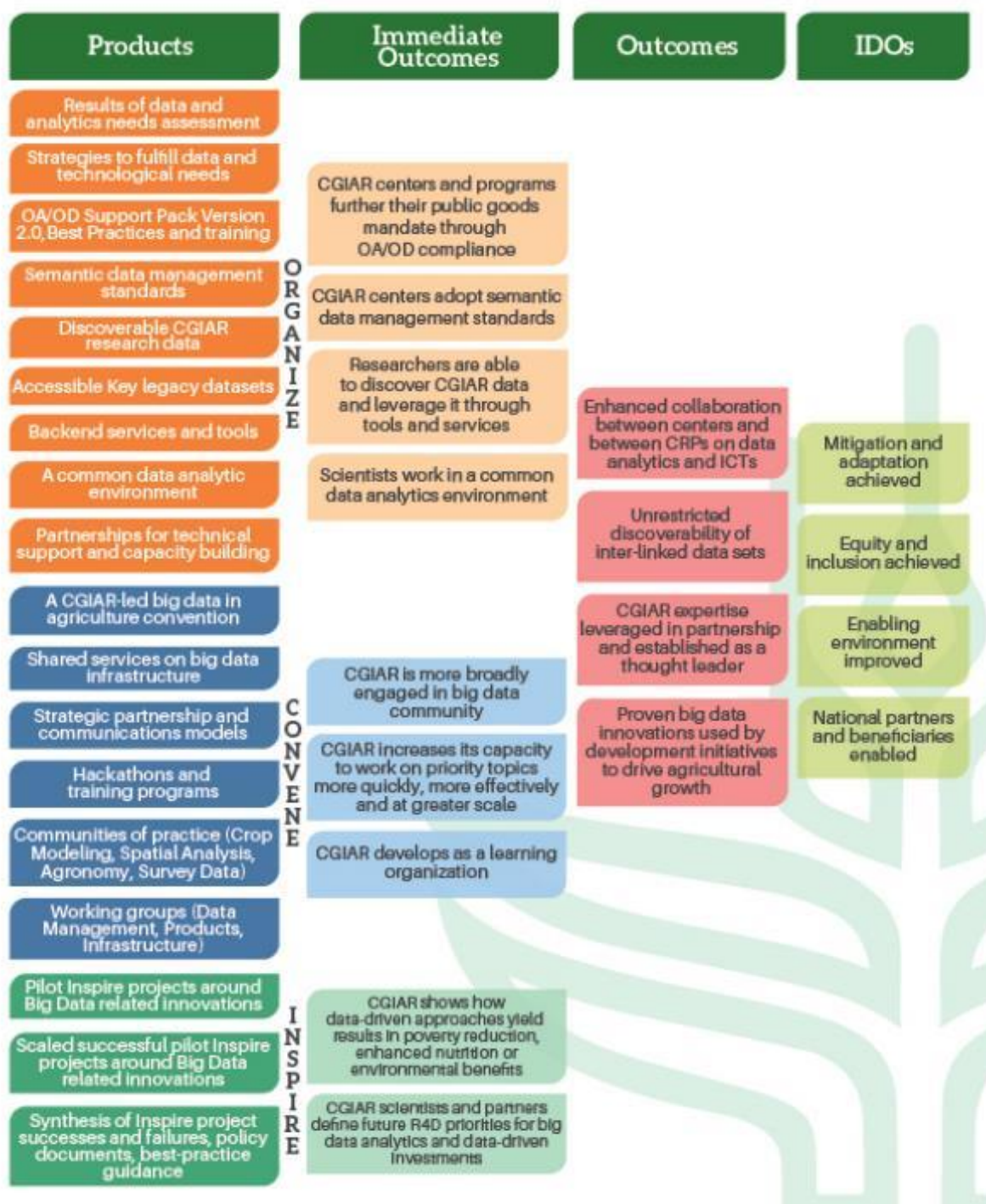


Figure 1 Impact pathways for the Big Data in Agriculture Platform

The primary assumption of this theory of change is that data are a valued commodity for development that can be harnessed to deliver growth in agriculture in developing countries. It also assumes that CGIAR and its partners can identify business opportunities where rural institutions are weak, to deliver benefits to marginalized, smallholder farmers. This assumption is central to the design of the Platform.

1.0.4 Capacity development

Analytics of big data is not yet a mainstream research approach in CGIAR. Therefore capacity development is pivotal to the Platform, to increase the ability of individuals to manage and use big data, and generate greater organisational capacity to embrace big data approaches to solving problems in CGIAR mission.

The Platform will address capacity development in all nine elements of CGIAR CapDev framework (CGIAR Consortium, 2015b), but focus in particular on five elements (the numbers are those assigned to the elements): (2) Design and delivery of innovative learning materials and approaches (primarily related to data management in Module 1); (5) Gender-sensitive approaches throughout capacity development (primarily related to targeted Inspire projects on this topic); (6) Institutional strengthening (across all Modules); (8) Organizational development (across all Modules); and (10) Capacity to Innovate (specifically through Module 3 Inspire). Wherever possible, the three Modules will link capacity development actions for greater effectiveness and efficiency, and work across Centres and CRPs. Capacity development is thus central to the impact pathways of the Platform as a whole, providing the mechanisms whereby increasing abilities to demand, undertake and utilize research lead to sustainable improvements in capacity to manage big data. Thus the Platform's strategy of capacity enhancement will address the full ToC cycle across the three Modules. It will be an inherent component of Module 1, Organize. Module 2, Convene is essentially a component of capacity development emphasizing elements (5) Gender-sensitive approaches, (6) Institutional strengthening and (8) Organizational development. The major projects in Module 3, Inspire, will be required to include capacity building of the partners involved and will directly contribute to enhanced capacity to innovate.

The Platform will stimulate the incorporation of state-of-the-art Big Data management and analytics into CGIAR research as well as develop several new tools and approaches for use within both CGIAR and partner institutions. While some of the tools are self-explanatory and will be easily integrated into existing research frameworks, others will require training in their use. To address this, the Platform will organize a series of capacity-building activities and develop the associated training materials.

The Platform will work at individual, organizational and institutional levels of capacity development, and with both researchers and research users. Because of its restricted focus and budget, the Platform will limit its downstream activities to work with boundary partners (Clark et al. 2011). The Platform will organize a series of training workshops on the use of, for example, big data management, data publication, and data visualization and analytical tools, ICT-based survey tools, and open source hardware platform. Training programs will target two levels of researchers and data scientists, introductory and advanced, and be opened to CGIAR partner organizations. The technology world is particularly male dominated, and so the Platform will take leadership in bridging the gap to build the capacity of women in particular during these training sessions.

1.0.5 Platform leadership, management and governance

The Platform proposes a governance structure that will have sufficient checks and balances to make sure that the Platform delivers the expected outputs and outcomes with high value for money.

Leadership and management: This Platform is co-led by CIAT and IFPRI, with both Centers sharing responsibility for the Platform to deliver. CIAT will take fiduciary responsibility and sign the performance contract with the system office. Both Centers have a strong track record on data-driven research. Leadership will be provided through a lean Secretariat, which will consist of a Big Data Coordinator, a communication and engagement specialist, and modest administrative support. The estimated annual budget for the secretariat is US\$250k. The Platform Coordinator will be hired through an open competitive process, and selected by a committee consisting of CIAT and IFPRI scientists. The Secretariat will be hosted by CIAT, and based in Cali where significant critical mass on Big Data already exists. 360 degree evaluations will be made of the Secretariat and its staff by all Platform participants

Module 1 will be led by Dr. Medha Devare who is currently in the Consortium Office leading open access and open data work. Her continuity in this role is important, given her knowledge, experience and established rapport with the data management and open access CoPs across Centers and CRPs. She will lead Module 1 either in her capacity as a member of the new System Office, or as an IFPRI direct hire. Details of this arrangement (e.g. hosting, reporting etc.) will be formalized once the System Office is established and dialogued with Platform Leadership.

To keep management costs down, Modules 2 and 3 will be led by the Platform Coordinator, who will oversee that the Communities of Practice (CoP) operate effectively, lead the organization of the Convention, negotiate and manage shared services (with support from a computer scientists position) and project manage the Inspire projects. The Platform Coordinator will be hired through an open, competitive process.

An Executive Management Team will meet bimonthly to actively manage the Platform, and this team will consist of Platform Coordinator, plus Module 1 leader, Dr. Andy Jarvis and Dr. Jawoo Koo, with other CoP leaders participating as observers.

Governance: The Platform proposes a Steering Committee (SC) to provide oversight and direction, and will formally report to the CIAT Board. As the signatory to the performance contract with the System Office, CIAT will have the overall fiduciary and operational responsibility for the Platform. The Steering Committee will be made up of 8 members, consisting of:

- One permanent member each from CIAT and IFPRI, designated by the Director General of each institution
- One permanent CGIAR System Office representative or the person they delegate
- One 2-year rotating CRP representative, nominated by the DGs and Science leaders group
- One 2-year rotating Center representative, nominated by the DGs and Science leaders group
- Three 2-year rotating partner representatives, nominated during the Annual Convention
- The Platform coordinator will also participate in the Steering Committee as an observer.

Rotating positions will be assigned to balance critical knowledge and experience on: Informatics, legal and intellectual assets, data management, data analytics. The SC Chair will liaise with the Directors-Generals of both CIAT and IFPRI as necessary. The workings of the SC will be reviewed after 24 months.

Programmatic decisions will be the responsibility of the SC. The SC will also consider how effective the Platform is in meeting its outcome and impact goals. The cost of the SC is estimated to be <\$20,000 per year.

An International Advisory Board (IAB) will also be set up, and will meet face-to-face once per year, and virtually once per year, to explicitly examine how the CGIAR Platform on Big Data connects effectively with other global and regional efforts. The IAB will contain 5 members, who represent major initiatives such as Open Harvest, GODAN, Open Data for Development amongst others. The IAB will focus its deliberations on the comparative advantage of the platform relative to other initiatives, and guaranteeing the relevance and novelty of the Big Data Platform with respect to the broader ecosystem of initiatives. The findings of the IAB will be reported to the CIAT Board once per year.

CIAT will be the main contracting body for the Platform and will sign the performance contract for the Platform with the System Office, and the sub-contracts with participating Centers/Partners. The CIAT Board Chair and Director-General will report to the System Office on the Platform as a whole. Ultimately, the CIAT Board will have authority over the entire Platform and its management but will also consult with IFPRI management as needed.

Detailed leadership and Management arrangement is provided in Annex 6.

1.1 PLATFORM BUDGET NARRATIVE

1.1.1 General Information

Platform Name	Big Data Platform
Platform Center	Lead CIAT

1.1.2 Summary

The total 6-year budget for this Platform is US\$30.2m, representing an annual budget which ranges from US\$3.6m to US\$6.6m. This represents approximately 0.5% of the total CGIAR budget. The logic of the budget is such that Module 1 receives greater budget in 2017 (68% total) and 2018 (58% total) to ensure that CGIAR data is managed effectively and achieves compliance of the OAODM policy. Module 2 consists of a US\$1.45m budget in 2017 which grows by a standard 5% each year. This Module maintains a fairly fixed cost as it is creating the enabling environment for Big Data efforts across CGIAR. The Module 3 Inspire budget is built in a way that it grows from Year 1 (US\$0.6m) to year 4 (US\$1.31m). As per the Theory of Change, CGIAR will first get its own data in order (Module 1) and then progressively grow Big Data analytics and research to demonstrate how data can drive agricultural growth (Module 3). A minimum of 15% of the total budget is earmarked for non-CGIAR partnerships, and all Centers are involved and will receive funds.

The Platform Secretariat has a cost of US\$300k. 40% of budget is considered to represent Capacity Building, and 17% is aligned to gender and youth related activities. A total of US\$600k is budgeted for impact assessment under Module 3.

The identified Funding Sources are primarily W1+2. Given that the Platform is a new initiative and a large portion of the funding is contributing to organizational strengthening, few bilaterals have been aligned at this stage, but it is expected that W3/Bilateral could grow in 2018 and beyond. The Platform proposes to align the current Gates Foundation grant to CGIAR to support Open Access/Open Data to this Platform. The Secretariat will be charged to fundraise greater Bilateral/W3 resources once established for specific components of the Platform. Opportunities for co-sponsorship of Inspire calls with the big data community will be aggressively sought and some initial conversations with partners and donors have already yielded significant interest.

Table 2 Overall platform budget

Module Name	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
Organize	4,336,320.51	3,172,574.31	2,261,673.74	1,159,962.09	1,125,489.82	1,192,411.46	13,248,431.93
Convene	1,455,300	1,516,077	1,579,603.14	1,646,517.85	1,716,339.85	1,789,267.18	9,703,105.02
Inspire	612,720	670,095.9	1,017,294.24	1,307,446.44	1,089,338.37	538,356.48	5,235,251.43
Management & Support	300,000	315,000	330,750	347,288	364,652	382,885	2,040,575
Strategic Communication	0	0	0	0	0	0	0
	6,704,340.51	5,673,747.21	5,189,321.12	4,461,214.38	4,295,820.04	3,902,920.12	30,227,363.38

Budget summary: Module 1 has the highest budget of all, and the main cost drivers include significant funding to centers to support them in OAODM policy compliance, in addition to funding non-CGIAR partners to bring new expertise to CGIAR data management. Module 2 costs are relatively fixed, and include basic funding of US\$100k to each CoP, plus US\$200k for the WG on Data products and data management and US\$200k for the establishment of shared services on priority infrastructural needs of the CRPs and Centers, including cloud storage, processing capacity and high-end analytical capacity. Module 3 budget progressively grows as additional Inspire projects are established, reaching a peak in 2020. Impact assessment of US\$158k per year is budgeted in years 2019 to 2022 to systematically assess Inspire project efficacy on poverty, nutrition and environment.

The Secretariat is budgeted at US\$300k with annual growth of 5%, and is counted as management and support cost. This represents less than 5% of the average annual cost of the platform.

Detailed summaries are provided under each module narrative summaries.

1.1.3 CRP Funding Plan

Table 3 Platform funding plan – funding needed

Funding Needed	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2	6,404,340.5	5,358,747.21	4,858,571.11	4,113,926.37	3,931,168.03	3,520,035.11	28,186,788.33
w3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	6,404,340.5	5,358,747.21	4,858,571.11	4,113,926.37	3,931,168.03	3,520,035.11	28,186,788.33

Funding Secured	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2(Assumed Secur	424,646	0	0	0	0	0	424,646
w3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	424,646	0	0	0	0	0	424,646

Funding Gap	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2	-5,979,694.5	-5,358,747.21	-4,858,571.11	-4,113,926.37	-3,931,168.03	-3,520,035.11	-27,762,142.33
w3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other3	0	0	0	0	0	0	0
	-5,979,694.5	-5,358,747.21	-4,858,571.11	-4,113,926.37	-3,931,168.03	-3,520,035.11	-27,762,142.33

Funding plan: As per the guidance call, the Platform is presenting this proposal without an assured W1+2 level of funding, hence the entire budget is marked as “needed”. Given that the Platform neither exists, nor has assurances that it will be established in 2017, bilaterals are not currently aligned to the platform. Nevertheless, W3 and bilaterals related to data management, data analytics and big data applications from Centres do exist, and can be aligned if and when the platform is established. The platform would target 1:1 matching of W1+2 with W3/bilaterals.

The Secretariat would also be charged with fundraising targets, especially through co-funding for Module 3 which is the most out-facing module which is potential attractive to donors and the private sector. Details for this are provided under Module 3 budget details in the Annex.

1.1.4 CRP Management and Support Costs

Platform management and support cost: The Platform Secretariat is budgeted at US\$300k in 2017, representing just 4.4% of total cost. This includes a lean 3 person team consisting of an internationally recruited platform coordinator, national staff level communications specialist, and a national staff level administrative assistant. Some US\$60k is budgeted for miscellaneous supplies and services, and an additional 30k for travel.

1.1.5 CRP Financial management principles

Platform financial management principles: Once established, the Platform Coordinator will have the responsibility to allocate W1+2 to the distinct modules based on the needs of users, funds availability and strategic direction. Decisions will be made in consultation with the Steering Committee. The Platform Coordinator will also oversee execution of the budget for Modules 2 and 3, hence will be fully responsible and accountable for tracking, reporting and revising Module 2 and 3 budgets. In the case of Module 1, IFPRI have ownership of the funds and will have full autonomy to decide upon the funds distribution. US\$50,000 is budgeted each year under Capital, and will be assigned based on a needs basis of all users across the platform.

1.1.6 Budgeted costs for key activities:

	Estimate annual average cost (USD)	Please describe main key activities for the applicable categories below, as described in the guidance for full proposal
Gender	450,000	Datasets made open access through Module 1 activities will be gender-disaggregated wherever applicable; this will be achieved in close engagement with CGIAR Community of Practice on Gender. Data standards developed in Module 1 will also include appropriate metadata for representing the gender dimension. Outcomes from Module 1 will help to democratize access to the information across gender. Elements of the Convention will focus on gender related topics from data management through to big data use in practice. Female participation in the convention will also be incentivized to break typical male dominated technology barriers down. Some Inspire project topics will be targeted specifically to address gender gaps, and innovate around big data and ICT means of reaching marginalized women farmers and female headed households. Additionally, Inspire projects on other topics will examine gender-related impacts of the innovation, and impact assessment studies will explicitly examine the gender-related impacts that big data innovations have.
Youth	350,000	We anticipate that increased access to agricultural data and ICT-based applications will help retain and attract more rural youth to agriculture. Young men and women will be considered a major target user group for the annual data/knowledge consultations and design of communication materials. Significant youth involvement from within the CGIAR and with partner organisations will be sought in both the Convention and in CoP to bring new ideas to the table. Similarly to gender, some Inspire project topics may specifically address youth related impacts. Given the technological nature of big data innovations, it is to be expected that youth are more likely to engage in projects given their additional capacity to manage ICTs. Likewise, impact assessment efforts will look at youth related impacts of different innovations.
Capacity development	1,954,000	Centers will receive dedicated funding for developing capacity to implement OA/OD and ensure science and data quality. It is anticipated that workshops will be also focus on or include capacity building as a core theme. The CoPs and WGs focus entirely on organizational Institutional Strengthening, and Organisational development, hence there is a heavy capacity building element in this module. Module 3 almost entirely focuses on building innovation capacity in CGIAR and with partners, hence a heavy load of budget can be considered aligned to capacity building.
Impact assessment	368,000	Consultations with the MEL CoP and data/knowledge CoPs will take place to assess the overall impact of OA/OD and define the scope for identifying opportunities and monitoring progress towards impact. Start-up funds for Centers to implement Altmetrics and ORCID will help in monitoring. Impact assessments of a range of big data related innovations for agricultural development are explicitly budgeted in years 3-6 under Module 3. Survey instruments and other techniques will be used in a range of countries and projects to systematically assess the impacts of scaled Inspire projects.
Intellectual asset management	163,000	Staff time of 0.66 FTE/year is budgeted to handle issues related to legal aspects of OA/OD and IA management, legal counsel, and contract/grant management. Some in-kind support on legal aspects of OA/OD and IA management will also be provided in-house for Center-specific issues. Intellectual assets resulting from Inspire projects will be managed as per IA guidelines by project leaders, with support from the Platform Secretariat. Some intellectual asset management will be considered under the CoPs and WGs as relevant.
Open access and data management	3,200,000	The entirety of Module 1 is aimed at addressing open access and data management across CGIAR. A significant part of projects in the Inspire module are expected to generate data and manage it in accordance with the OAODM policy of CGIAR. A significant focus of both the Convention and the CoPs/WGs is around open access and data management.

Communication	320,000	A large component of the Convention is focused on innovative communication around the topic of Big Data and Agricultural Development. Development of advocacy and training materials, support for open agricultural data publications, and design/printing of communication materials are included. Significant communication activities will be made around Inspire projects from the Secretariat, but also by project leaders themselves, in order to disseminate good practice in big data related to agricultural development.
----------------------	---------	--

2 BUSINESS CASE

2.1 Expertise and track record

CIAT:

CIAT's data science predates the concept of big data.

Data as IPGs: In the 1970's, CIAT constructed a climate database of 1000s of weather stations across the tropics. It identified priorities, set geographical targets for crop improvement, and tailored solutions to site-specific problems. The dataset fed into the [WorldClim](#) suite of climate surfaces, which was published online. It has been downloaded hundreds of thousands of times. The WorldClim paper is the second-most cited paper in CGIAR's history. This founded a culture in CIAT that values high-quality data for improved decision making as an IPG. It promotes the principles of open access and invests long-term in development of datasets and data infrastructure.

Analytics Tools: CIAT hosts data resources that help CGIAR achieve its strategic goals. These include [CCAFS Climate](#) (with 50,000 users), [AgTrials](#) (database of 35,000 agricultural trials from CGIAR and partner institution), [RTBMaps](#), tools such as [Climate Analogues](#) (analytics of comparable climate conditions globally), [MarkSimGCM](#) (daily weather generator), and ICT platforms (e.g., mobile applications for farmer field data registries, analysis and delivery of recommendations).

Recognitions: CIAT won the [UN Global Pulse Big Data and the Climate Challenge Prize](#) (2014) and the [Ebbe Nielsen award for innovative research in bioinformatics and biosystematics](#) (2009). RTBMaps won the Special Achievement in GIS Award (2013) and the ComputerWorld Data + Editor's Choice Award (2014).

IFPRI

IFPRI is an inter-disciplinary thought leader applying socio-economic analysis to support formulating agricultural policies. It has a long history of interdisciplinary policy analysis utilizing harmonized, large spatial datasets across scales and modeling capacity. Headquartered in Washington, DC, with many big data technology partners in proximity, IFPRI has a comparative advantage to coordinate convening activities. IFPRI's data analytics-relevant capacity includes:

Strategic Foresight: Using the IMPACT model, IFPRI leads the global-scale analyses of food supply and demand, trade, prices, and food and water security with technology and global change scenarios at sub-national level. Model outputs provide foresights into the policies and actions needed to feed the world, reduce poverty, and protect the natural resource base, all relevant to CGIAR SRF and SDG Goals.

Analytics on Grids: In 2008, IFPRI pioneered the development of global-scale gridded crop modeling framework using DSSAT suite of crop models on a high-performance computing cluster. The framework has been further developed for use in global and regional-scale ex-ante studies of impact assessment. IFPRI supports installation of the framework at other CGIAR Centers.

Data as IPGs: IFPRI publishes harmonized, spatial databases of multi-disciplinary data layers on 10 km global grids. The database provides 700+ indicators, with production statistics for 42 crops. This dataset is a community-driven knowledge product with participation of all CGIAR Centers, and is used as the baseline for global and regional-scale studies. IFPRI also publishes ASTI, which tracks agricultural

research and development in 80+ countries. ASTI supports science and technology policymaking and builds national and regional capacity for data collection and analysis.

PLATFORM PARTNERS

All 15 Centers and 12 CRPs are also partners. Additionally, the Platform has engaged with 47 external partners ranging international institutions, universities to private companies. They cover public to private, developing to developed country, and analytics to ICT deployment. Amongst these are major global players on big data analytics such as IBM, big data infrastructure expertise such as Google, Amazon, through academic institutions such as MSU, U. of Florida, King's College London, and Big Data start-up and investment companies such as Bayes Impact, DataPop Alliance. Details of these partners is available in Table 1.

COMPOSITION, EXPERTISE AND EXPERIENCE OF PROGRAM TEAM	
AkiraChix	African Technology Innovation hub, providing capacity building and access to a network of women technological entrepreneurs
Amazon	Amazon Web Service (AWS) provides a suite of cloud-based computation tools and application development environment. CGIAR's datasets are also served.
Arizona State University	President of the University Consortium for Geographic Information Science (UCGIS), spurring collaboration between UCGIS and CGIAR, and furthering the role of big data in decision science
ASARECA	Promoting adoption of agricultural technologies in 11 member states in East and Central Africa. Collaborates with IFPRI/PIM/HarvestChoice on technology mapping and targeting.
Battelle Memorial Inst.	Battelle's researchers use advanced analytics and machine learning to help companies track and monitor relevant issues in both structured and unstructured big data sources.
Bayes Impact	Bayes Impact brings a team of top Silicon Valley data scientists and engineers who build data-driven software applications that can optimize agricultural outcomes at scale
CABI - GODAN	Open data in agriculture, and linkage to GODAN initiative
CEO of Novogene Bioinformatics Technology Co. Ltd	To sequence all the collection of cassava lines and common bean lines and to work together to develop the cutting-edge data analysis and bioinformatics technologies, and to use Pan-genome approach to investigate the detailed genetic diversity at population level
Columbia University	CIESIN specializes in online data and information management, spatial data integration, and interdisciplinary research. IFPRI collaborates with CIESIN for various spatial analysis products.
CSIRO	Agriculture Flagship develops new tools and management practices, including crop models, IPM, and precision agriculture, to increase profitability and productivity.
Data-Pop Alliance	Connecting hub, alliance ensuring that data and tools serve the interests of people across the globe, catalyzing and coordinating developments and innovations in the use of Big Data to help serve the cause of human progress
Disney Research	Disney Research uses UVAs and 3D sensors from UAVs to model vegetation and computer vision to research human interaction and logistics such as crowd counting.
FAO	AIMS team collaborates with CGIAR Consortium Office, IFPRI, Bioversity, and other Centers to support co-developing ontologies, common vocabularies, and knowledge management tools.
First Mile Geo	FirstMileGeo's platform streamlines the design of questionnaire, data collection using mobile devices, and visualizing and analyzing the data online.
GBIF	Global Biodiversity Info. Facility provides access to various types of biodiversity-relevant datasets. GBIF collaborates with Bioversity.
GeoPoll	A mobile polling platform that help organizations to conduct surveys through the mobile phone. IFPRI uses GeoPoll to survey maize farmers' technology use in Tanzania.
Global Crop Diversity Trust	Input to the Digital Genebank Inspire Project, linkage to DivSeek and scaling out concepts to other genebanks
Google	Earth Engine provides petabytes of public remote sensing data and a library of analysis functions to perform complex analyses in Google's cloud computing environment.
IBM	Transforming industries and professions with data; Remaking enterprise IT for the era of cloud computing; Reimagining work through mobile and social technologies; Rethinking the challenge of security.

IIASA	Ecosystems Services and Management Program researches to improve our understanding of ecosystems and monitoring their current state and ecological thresholds.
King's College London	Partnering with CGIAR to generate policy support systems on water, ecosystem services.
McKinsey & Co.	Agricultural Commodity Research Engine (ACRE) delivering data analytics and decision-support tools to rapidly test and provide feedback on geospatial data along the value-chain and targeted insights to emerging markets.
Michigan State Univ.	PhotosynQ Project: Expertise on developing and providing low cost, hand-held measurement device to collect and share plant health data.
Michigan State University	Generating user centered tools to expand the base of data users and expand the reach of big data into decision making.
NIAB	Genome sequencing and collaboration on Digital Genebank ideas.
Oak Ridge National Lab	Geographic Information Science and Technology group uses supercomputers and geo-computational methods to solve a wide variety of national and global problems.
Penn State University	PlantVillage: using smartphones and smart crowds for food security
Purdue	Big Data collection and analytics as it applies to plant improvement – sensor development and data collection innovation (robotics, UAV, etc.), field-scale phenomics/plant analysis, genome to phenome data integration.
Rothamsted	Collaboration on site-specific agriculture approaches, linkage with UK Agritech initiative.
Royal Holloway University of London	Generation of large non-targeted and targeted metabolomics datasets augmenting the chemotyping of diversity germplasm and supporting breeding programs across the RTB portfolio.
SpatialDev	SpatialDev has years of experience working with CGIAR and many agricultural research and development partners to develop web-based data visualization and analysis tools.
Stanford University	FSE (Center for Food Security and the Environment) develops satellite imagery-based mapping of crop production and performances. CIMMYT and IFPRI collaborates with FSE.
SUPAGRO	Big data and AgroTICS
UC Davis	The Geospatial and Farming Systems Research Consortium (GFRC) supports the adoption of sustainable intensification, using open data, analysis tools and provides training to partners.
UN Global Pulse	Advisory support for tailoring of technical solutions to ensure utility for United Nations agencies and other stakeholders.
UNESCO-IHE	Leads the development of earth observation technologies for soil, land surface and atmospheric processes, with a specialization in agricultural water management.
University of Applied Sciences of Western Switzerland (HEIG-VD)	Intelligent data analysis, providing practical solutions based on machine learning analytical tools to deal with real-life challenges in agriculture.
University of Florida	Building distributed or harmonized databases to support agricultural research for development, and developing ICT tools that facilitate interoperability of data, models, and other knowledge products.

University of Nebraska	Water for Food Institute uses remote sensing applications for irrigated agriculture, hydrology and natural resources monitoring, including applications using low-cost airborne sensors.
Washington State Univ.	Leads the development of DSSAT, a crop systems modeling platform, and coordinate training programs. Collaborates with Global Futures, CCAFS, PIM, and HarvestChoice initiatives.
WUR	Big data and interoperability issues
World Bank, Agriculture	Supporting scaling up and out the data driven agronomy experience from LAC to Africa and East Asia and the Pacific regions.

Table 1 Platform partners

2.2 Access to Materials, Sites and Services

Unimpeded access to Center and CRP datasets, materials, sites, services and networks will be a pre-requisite for the Platform to succeed. Wide access is more likely if it occurs via a bottom-up rather than top-down process. The central idea of the Convention (see Communication Strategy, below and **Module 2, Convene**) is to bring together relevant people across Centers/CRPs and collaborators to create consensus on open access and sharing. In this way, the curators and managers of data resources will work together to address concerns and overcome resistance to sharing materials. Module 2 will also negotiate long-standing service provision for hosting major CGIAR datasets and ensure their availability to users with the highest standards.

2.3 Interactions between the Platform and Users

One of the primary objectives of the Platform is to bring together CGIAR Center and CRP data in such a way that they are available to end users, ranging from policy makers to smallholder farmers. It is therefore axiomatic that these end users must be involved in developing the tools to access the data. The direct users are therefore scientists in CRPs and their partners.

The key communication space between the Platform and primary Center/CRP scientists and partners will be the annual Convention. This will not be the typical big conference with keynote speakers and individual presentations, but a more engaging format where the participants are encouraged to interact together to address a suite of common themes. The Conventions will be patterned on the successful meetings of a successful informal group in CGIAR called the [Consortium for Spatial Information](#).

Network mapping and connectivity mapping with some reportable outcomes will be mainstreamed in the Platform to ensure relevance. Too often networks are discussed as a soft objective with convening's and untraceable interactions driving that objective. This Platform will track the network, map its' nodes, hubs and connections and be able to analyze quantitatively the outcomes of this objective. This network mapping also serves to allow navigation of the network from the outside by connecting expertise, activity, and communities within the network to stakeholders. Web tools to allow navigation and engagement with the network will increase CGIAR visibility to external audiences as well as increase the velocity of internal collaborations.

The Platform will also aggressively pursue active ideation around the creation of advanced analytical tools that are accessible. For example, GIS was exclusively the province of trained geographers and analysts until recently. Advances in the technology have made it more useable and drastically shortened the learning curve, hence GIS has become democratized and currently has a much broader role in supporting decisions than ever before. Opportunities exist in this vein with statistical analysis, landscape analysis, and climate measurement and in resilience.

The partners to this Platform include both upstream knowledge generators, through downstream knowledge users. Collaboration with experts in this field, such as Google and IBM, who have indicated their interest in the Platform, will provide the expertise and insights needed. Details of the committed collaborators are in Annex 4.

2.4 Intellectual Asset Management

The SC will be responsible for the management of intellectual assets of the Platform. In particular, the SC will ensure that the Platform complies with CGIAR Principles on the Management of Intellectual Assets, which deal with the dissemination of intellectual assets for maximizing global accessibility and impact. The Platform's Intellectual Assets (IAs) will include data and databases, analysis tools, publications and other information products as defined by article 3 of the CGIAR Open Access and Data Management policy. All products produced by the Platform will be disseminated using open access principles, with clear branding to recognize those responsible for producing the IA. The products will have suitable open licenses to ensure maximum possible access and reusability. In the cases where particular copyrights apply (for example, in the case of some high impact journal articles), the Platform will commit to cover the fees to make the articles open and reusable.

When working with private sector entities, the Platform will clarify that it is committed to open access on knowledge products and will abide by any rules placed on the partnership, but final products will be made public in accordance with the agreements, subject only to confidentiality obligations as may be associated with restrictions permitted under article 6 of the CGIAR IA principles. The Platform will ascertain that data collected about or on human subjects (human beings or groups of human beings) has passed the ethical review put in place by the relevant CGIAR center or other partner organization.

The Platform will provide a checklist regarding ethics in social research as part of a general ethics in data collecting protocol. Bringing resource poor farmers to Big Data and Big Data to resource poor farmers requires special attention for privacy and cyber security issues. The CGIAR has a special responsibility in protecting the privacy of the resource poor beyond the legal minimum.

The proposal does not identify non-standard pathways of dissemination such as those involving legal restrictions or regulations, third-party technologies or innovative approaches. This is because it is not possible at this proposal stage to foresee what these might be. Indeed, one of the Platform's objectives is to create novel pathways to make CGIAR data more readily available and useful to its clients, resource-poor farmers. One of the SC's responsibilities will be to seek ways to create non-standard pathways on an ongoing basis. All three modules are dedicated to the objective of creating new ways to make CGIAR information available to end users. Platform partners will be key players in developing these novel approaches and evaluating these possibilities during implementation of the Platform.

2.5 Open Access Management

Open access management is central to this Platform, especially under Module 1, with the objective of supporting CGIAR to comply with OA/OD principles. In keeping with CGIAR Open Access and Data Management Policy (OADMP), research data, tools, and associated information generated under the Platform will be made available for indexing and interlinking, so that research outputs are open via FAIR principles (Findable, Accessible, Interoperable, Re-usable) to enhance innovation, impact, and uptake. The OADMP complements CGIAR Principles on the Management of Intellectual Assets.

Open access management across Centers and CRPs includes consistent operationalization of the OADMP and providing the overall framework, systems and structure. This will be operationalized in **Module 1, Organize**; research outputs hosted within the existing Platforms within CGIAR (e.g. CGspace, Dataverse, CCAFS climate) and external Platforms will comply with OA/OD requirements and FAIR principles.

A key performance target for this Platform is that it supports CRPs and Centers to comply with OADMP through a range of support mechanisms from capacity building, through targeted technical support to generation of new tools, services, portals and web platforms that ensure that CGIAR data is truly open access.

The Platform will deal only with information about germplasm provided by the Centers and CRPs, not with germplasm itself. Compliance with the International Treaty of Plant Genetic Resources for Food and Agriculture (ITPGRFA, www.planttreaty.org) and its Nagoya Protocol is the responsibility of the Centers and CRPs that deal with germplasm. The SC will be charged to ensure that the Centers and CRPs certify that the information they provide to the Platform and its partners is in compliance with the ITPGRFA and its Nagoya Protocol and promotes farmers' rights.

2.6 Communication Strategy

Strategic communications offer a set of powerful tools and approaches that can contribute to the Platform's outcomes and CGIAR's SRF. Communications activities are aligned with the Platform's impact pathways and are an essential component of the ToC. They are designed to deliver on the Platform's overall goals and a range of measurable communications objectives that promote the science of big data. They will inform major initiatives and policies on big data, provide open access to knowledge, and encourage learning and sharing of information across the Platform and its partners.

The Platform will employ a suite of well-established tools and approaches (e.g. through website with blog posts, micro-blogs on twitter, linkedIn pulse blogs, targeted messaging through local national and regional). These will deliver tailored messages to program partners and next users. Communications will be custom targeted at end users, ranging from scientists to policy makers to smallholder farmers. The Platform will also draw upon its multiple partners to leverage the diffusion and interaction with its members and audiences. The platform will engage with specialized media, i.e. in the area of data analytics, and ICT4D.

Innovative communication approaches such as Hackathons and Killer App competitions and prizes will also be developed around the Convention to increase visibility of the Platform as a go-to-place for knowledge and insights on big data in agricultural development. Further, these kinds of events will be used to draw additional partners to the table, and achieve specific Platform goals through concerted crowdsourced efforts to innovate and build development solutions.

Communications will be supported through budget allocations and staff time from a variety of sources, including the Secretariat, Platform Modules, Platform-financed projects, and Centers/CRPs and partners. Collective action and collaboration will be essential to deliver impactful communications.

2.7 Risk Management

The Platform will maintain a risk catalogue following the approach used by CIAT and CCAFS, which will be updated once a year by the SC at its annual in-person meeting. The risk catalogue will include details on each of the risks, their likelihood and impact, mitigating measures and risk owner. The results of the risk assessment will be reported each year to the CIAT Board and will be discussed by the Platform Coordinator with the Chair of the CIAT Board's Audit and Risk Committee. The Platform Coordinator will convey any suggestions to the SC for discussion and implementation. An initial appraisal of some (not all) risks are nevertheless provided below.

One particular risk is that there is insufficient buy-in from CGIAR leadership (CRP directors, Center directors, System Office) to pay sufficient attention to compliance of the OADMP policy. The Secretariat will very strongly engage with CGIAR management and leadership to ensure that this topic

is prioritized and given sufficient attention. Module 1 will provide incentives (technical, financial and social) for Centers and CRPs to achieve OA/OD compliance.

Loss of data is a risk that CGIAR faces today and this risk will only increase in the future as data becomes a more and more important commodity, and so this Platform will address this risk by ensuring adequate backup, and through negotiation of shared service agreements for cloud based backup of essential CGIAR data resources.

Whenever working with data that includes information about individuals, there are privacy related risks. Module 1 will work with Centers and CRPs to ensure responsible management of data with respects to privacy and support CGIAR to have strict guidelines about anonymity, privacy and confidentiality in all types of data that the organization manages. Having an ethics in research protocol and abiding by it will minimize the probability of privacy related conflicts occurring. Privacy risks which can damage reputations, and create a lack of trust for future data-related activities with those individuals. There are also legal and liability risks associated thereof, since a violation to the individual's privacy is sanctioned by a considerable number of countries.

Making data open access without the ability to show compliance with ethics in research principles and protection of privacy can legally expose the institutions that make such data available.

3. MODULES

3.1 MODULE 1: ORGANIZE

3.1.1 Modules Narrative

3.1.1.1 Rationale and Scope

RATIONALE

The CGIAR pursues multiple pathways for impact through the CRPs and SRFs. The Platform held scoping consultations with CGIAR researchers and partners, which identified key requirements and components of the enabling environment to realize pathways to data-driven impacts, as well as current gaps in capacity. This Module addresses organizational and technical environments that will enable CGIAR to embrace the power of big data analytics. This will occur within following three high-level domains.

Domain #1: CGIAR's Open Access and Open Data Compliancy

New techniques in data science (such as machine learning) offer scope to complement the existing science toolbox of international agricultural research. The Platform will implement and capitalize on these technologies, which will provide opportunities for interoperability between data, tools and models. CGIAR's diverse range of research datasets will then be fully open and have a common meaning. This will require CGIAR-wide transformation and the appropriate supporting infrastructure. The Consortium Office has been implementing the CGIAR Open Access and Open Data (OA/OD) policy to address organizational and technical challenges. Its primary focus was to foster a change in culture and institutional incentives, along with technical and human infrastructure. The initiative's first phase assessed the infrastructure and capacity needs, developed a framework to prioritize legacy data, and provided coordinated support to Centers and CRPs. It also developed a framework to assess the initiative's impact. The initiative made progress, but needs additional investment and effort to meet its goal of seamless and consistent information discoverability, integration, interoperability and reuse. The OA/OD Phase II proposal, submitted in December 2015, seeks to strengthen the volume, variety, and quality of openly-accessible research outputs from CGIAR and its partners. These objectives are

closely aligned with the objectives and expected outcomes of this Platform. We therefore incorporate the OA/OD Phase II into this Module.

Domain #2: Provision of Enabling Datasets, Tools, and Services

CGIAR's primary strength lies in generating and managing high-quality global research for development. CGIAR does not, however, yet have a comparative advantage in terms of the technical capacity and resources required to develop big data analytics tools and supporting services. Participants in the scoping consultations emphasized the need to access third-party key datasets, tools, and supporting services. A selected set of these will be made available through the Platform, prioritized based on the needs of CRPs and the amount of resources available.

Domain #3: CGIAR-wide Adoption of a Common Data Analytics Environment

Although online tools help users to discover and explore datasets easily, more advanced analysis may still be required locally after the dataset (or its subset) is downloaded. By promoting a common data analytics environment for all researchers to use as a standard research tool, we anticipate lowering the barrier of practical use of big data analytics in CGIAR. This does not imply any enforced exclusivity, rather researchers with advanced skills will still have their choice of tools as they fit their needs. Yet, wide adoption of a common analytics environment will help provide focused capacity building activities, strengthening consortium-wide analytical capacity in a more collaborative way. It will also broaden the use of big data analytics across the Centers and CRPs.

3.1.1.2 Objectives and targets

Objective #1. Assess the Data and Technology Needs

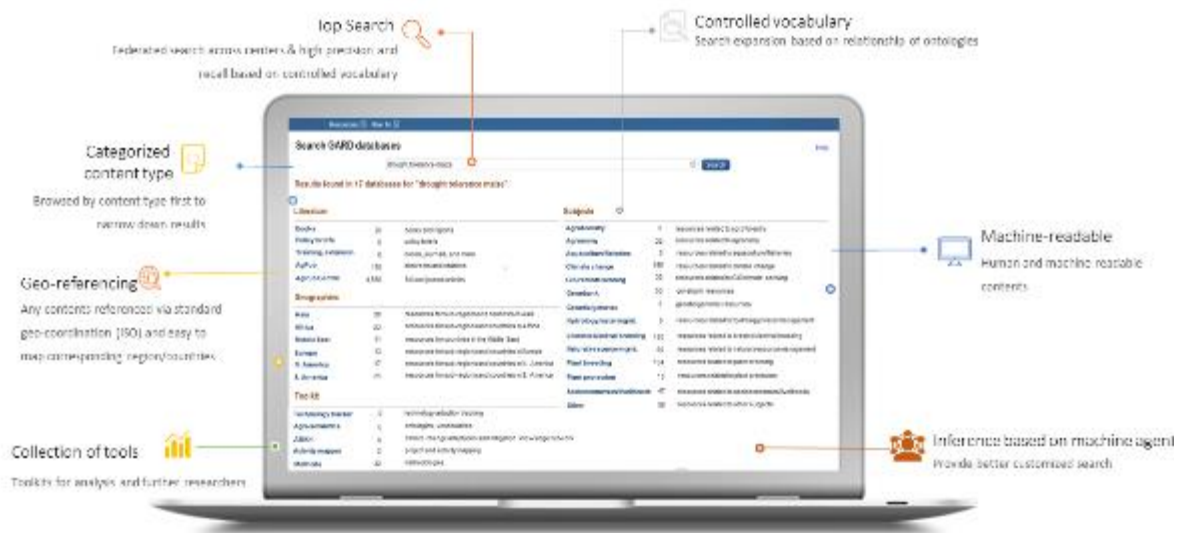
In collaboration with partners in public and sectors in selected focus countries, a thorough data landscape analysis will be conducted to understand their potential *data and analytics* needs and current gaps. The analysis will also define potential business cases where new data-driven insights could help raise investments in agriculture and guide business decisions, from product innovation to market access and infrastructure development. This assessment will be iteratively *implemented in regular basis* to keep the Platform's relevance for the newly emerging agricultural research and development topics and big data technologies and to help monitoring the Platform's impact.

MODULE 1. ORGANIZE	
OBJECTIVE 1. ASSESS THE DATA AND TECHNOLOGY NEEDS (10%)	
Target 1.1. Assess data and analytics needs for impact First stage of the needs assessment will focus on the identification of strategic opportunities for utilizing big data analytics within CGIAR SRF. This assessment will be highly structured to ensure measurable impacts from the data-driven insights, rather than to fulfill the reporting needs on, for example, the use of data. The Platform will closely collaborate with a well-qualified external expertise to provide guidance in the process. An Impact Assessment framework will be also developed in collaboration with CGIAR's Monitoring, Evaluation, and Learning CoP. The process will also include stakeholder consultations in priority countries.	
Output 1.1.1.	Technical consultation to assess data and analytics need for realizing CGIAR's impact pathway and monitoring its progress toward the SRF Goals.
Output 1.1.2.	Regional stakeholder meetings (Africa, Asia, and Latin America and Caribbean) to understand data and analytics needs to help achieving their development goals.
Output 1.1.3.	Consultation with CGIAR's Monitoring, Evaluation, and Learning CoP to develop and implement a robust framework to monitor progress and assess impact.
Target 1.2. Identify the data and technological solutions to fulfill the needs Juxtaposing the needs assessment and the technical capacity across CGIAR and its partners, in-depth consultation will be organized with key partners to help strategize to effectively fulfill technical needs with available resources. Given the rapid technical progress in the big data analytics landscape, this assessment will be revisited through the course of the Platform on a regular basis and in agile ways.	
Output 1.2.1.	Annual landscape analysis on big data technology to review promising solutions that can best fulfill the needs

Objective #2. Open CGIAR Research Outputs for Advanced Utility

To harness the power of big data analytics, a priority requirement is for research data to be easily found, well annotated, and readily integrated and interoperable for further utility. Activities will be implemented to make all CGIAR Centers and programs OA/OD compliant, all Centers to adopt interoperable data management standards, and make all CGIAR research data discoverable and utilized to power real-time decision support systems featuring advanced artificial intelligence technologies and natural language-based search interfaces.

MODULE 1. ORGANIZE	
OBJECTIVE 2. OPEN CGIAR RESEARCH OUTPUTS FOR ADVANCED UTILITY (30%)	
Target 2.1. All CGIAR Centers and programs to become OA/OD compliant Various formats of materials will be developed to raise broad awareness of OA/OD and to guide researchers on best practices, workflow, and FAQ on licensing, privacy, and ethics issues. Training and assistance will be provided to operationalize OA/OD and address key issues such as data quality and interoperability.	
Output 2.1.1.	An expanded and enhanced OA/OD Support Pack Version 2.0, with materials and presentations and consistent messaging designed to help increase awareness among and support for researchers across Centers.
Output 2.1.2.	Best practices for embedding OA/OD into researcher and Center roles and responsibilities.
Output 2.1.3.	Strengthened internal and partner capacity in support of OA/OD via training and professional development.
Target 2.2. All CGIAR Centers to adopt interoperable data management standards Semantic interoperability is the key to enable unambiguous data exchange, further allowing machine-computable logic, inference, knowledge discovery, and data federation. CG Core Metadata and Agri-Semantics Standards (e.g., Crop Ontology, Agronomy Ontology) will be completed, refined and deployed across CGIAR repositories. Additional domain-specific technical specifications (e.g., common spatio-temporal reference system) will be also developed as necessary.	
Output 2.2.1.	Semantically interoperable results enable interlinkages among research outputs.
Output 2.2.2.	Interoperable, harmonized datasets and technologies.
Target 2.3. All CGIAR research data are discoverable CGIAR Centers and CRPs maintain data and publication repositories. Many are indexed by search engines, yet some still represent less-discoverable silos. A new CGIAR Open Data Exchange will be developed and launched, by harvesting research outputs and metadata from distributed repositories and other relevant Platforms. Through a user-friendly natural language-based data discovery interface implemented in multiple languages, users will be able to access multiple sources to combine, discover, and share datasets, and utilize them in analytical and visualization tools.	
Output 2.3.1.	Seamless discovery of and access to research outputs from multiple sources.
Output 2.3.2.	Enhanced CGIAR data repository infrastructure.
Output 2.3.3.	Implement APIs to allow querying data in natural language from multiple databases (e.g., What crops are grown in Northern Ghana?) in multiple languages (tentatively in four languages: English, French, Spanish, and Swahili).
Output 2.3.4.	Enhanced search interface to query data in natural language with voice-recognition and geolocation of the point of query (i.e., Siri for Agriculture).



Objective #3. Provision of Enabling Datasets, Tools, and Services

MODULE 1. ORGANIZE**OBJECTIVE 3. PROVISION OF ENABLING DATASETS, TOOLS, AND SERVICES (30%)****Target 3.1. Key datasets identified and published**

In-depth feasibility and availability reviewed soon after the Platform's launch, a collective decision will be made on strategizing to make these data available, starting with a few key CGIAR use-cases to ensure that datasets are relevant for the purposes of targeting research and performance monitoring.

Output 3.1.1.	Decisions made on which key datasets to invest in as the Platform's flagship products, through a prioritization framework and consultations across CGIAR
Output 3.1.2.	Dataset providers identified (joint-partnership between CGIAR-internal and external parties) and agreements developed to ensure relevancy and quality
Output 3.1.3.	Key datasets made available through the Data Exchange, promoted and utilized across CGIAR and partners
Output 3.1.4.	Annual review of the investment conducted and decisions on continuation (or adjustments) made for the following years

Target 3.2. A suite of useful backend services and tool identified and launched

From the scoping consultations, a number of services and tools were identified as useful and enabling. A selection of the identified services will be provided through the Platform.

Output 3.2.1.	Decisions made on which key backend services and tools to contract, through prioritization of use cases and consultations across CGIAR Centers and programs
Output 3.2.2.	Small-scale fixed-term service agreement developed with a pilot group in CGIAR to evaluate as early adopters, provide feedback, and promote.
Output 3.2.3.	Annual review of the service conducted and decisions made regarding service expansion to scale, or adjustments on scope.

EXAMPLES OF ENABLING DATA AND TOOLS	
Suggested at the scoping consultation meetings (24 Feb – 3 Mar, 2016)	
Useful Datasets	Identified Expertise in Partnership
Global spatially-explicit agro-ecological, bio-physical and socio-economic data to help better understand and contextualize complexity in agro-food systems (e.g., mapping of crop geography, population and demographic, health and nutrition, poverty and wellbeing, infrastructure, technology adoption, and market accessibility)	IFPRI, ILRI, ICRAF, ICARDA, IWMI, CIAT, CIMMYT, Stanford University, Columbia University
Global macro-level food security and global development data for monitoring the progress toward the achievement of sustainable development goals and related CGIAR goals over time (i.e., SDGs, SLOs, IDOs, and sub-IDOs)	FAO, IWMI, World Bank, Columbia University
Useful Third-Party Backend Services	Identified Potential Service Providers
Network of sensors to measure plant physiology, weather, and soil	PhotosynQ, TAHMO
Coordinated survey data collection and management platform	FirstMileGeo, SurveyCTO, Fulcrum
ICT-based rapid data collection platform	GeoPoll
Cloud-based drone-captured imagery sharing and analysis platform	DroneDeploy, SpatialDev
Cloud-based computing environment	Amazon, IBM
Cloud-based remote sensing imagery analysis platform	Google, DigitalGlobe
Interactive data visualization	Tableau, SpatialDev

Objective #4. Implementation of a Common Data Analytics Environment

Success of the Platform will further accelerate when researchers across CGIAR and partner institutions use an interoperable data analytics environment and inspire each other. Although online tools will help discoverability, visualization, and simple analyses of datasets, more advanced analytics will be performed locally after the dataset (or its subset) is downloaded. By promoting a common, open data analytics environment, we anticipate lowering the barrier to practical use of big data. The Platform will also establish a code repository housing codes from various studies across CGIAR, and allowing researchers to explore practical examples from peers and adopt big data analytics in their work. Shared analysis codes will also help accelerate customization of analytics and avoid possible duplications. This will ultimately contribute to the advances in science by enabling reproducibility of published work and increased efficiencies.

MODULE 1. ORGANIZE	
OBJECTIVE 4. IMPLEMENTATION OF A COMMON DATA ANALYTICS ENVIRONMENT (30%)	
Target 4.1. A common data analytic environment selected and promoted Through technical consultations and survey across CGIAR and partners, a common data analytics environment will be selected and promoted. Mechanisms for providing technical support will also be established to facilitate its accelerated adoption.	
Output 4.1.1.	CGIAR's common data analytics environment selected through technical consultations
Output 4.1.2.	Technical support mechanism established
Output 4.1.3.	Best practices and analytical methods and tools identified, developed/adapted, made accessible and promoted.
Target 4.2. Increased use of big data analytics in CGIAR Collaborating with academic institution partners identified as champion users of the selected analytics environment, virtual and on-site trainings will be provided to adopt best practices in developing data analysis, data management, and managing codes and documentation.	
Output 4.2.1.	Established partnerships for technical support and capacity building
Output 4.2.2.	Virtual and on-site training and capacity building activities provided
Output 4.2.3.	Code repositories established for internal and public sharing of best practices and exemplar codes from peer researchers

3.1.1.3 Science quality

An implemented OA/OD strategy will generate accessible and ready-to-use data for analytics and research. Varying data quality is likely to impinge both on insights generated from the use of analytics and on the ultimate outputs. Unlike publications, it is not easy to implement peer-review of datasets. Nevertheless, robust scientific review will be useful, but it is impractical to expect reviewers to be accountable for the quality of large multi-disciplinary datasets. Depending on the characteristics of data products, whether they are used as inputs to other analyses (e.g., real-time field measurement data-stream), analytics outputs (e.g., forecasted probability of crop failure), or synthesized data-driven actionable insights (e.g., policy recommendation to prioritize investment options), the Module will establish a Science Quality Assurance Strategy as a part of the annotation of knowledge products.

Strategy #1: Provide data quality training and develop quality control protocols

To ensure the quality of big data used in the analytics, the Module will develop a set of protocols to provide guidance for the Platform partners and contributors. The protocols will assist to assess the quality of datasets, especially where there is no associated peer-reviewed publications, and identify potential issues with them. User feedback will be used as a proxy to monitor the quality of datasets, which is crucial to improve their quality.

Strategy #2: Convey the probabilistic nature of data analytics outputs with associated uncertainties

While researchers attempt to answer specific research questions, by definition big data analytics generate probabilistic outputs. This is particularly the case for any empirical or model-derived insight, including agricultural risk (e.g., 70% chance of doubling crop yield by doubling the amount of fertilizer applications depending on rainfall pattern). Conveying the concept of risk to smallholder stakeholders

may pose a communication challenge, but communication specialists will help avoid misinterpretation of analytic outputs. Where applicable, estimates from multiple sources using different approaches (e.g., disaggregated data of crop production) will allow diagnostic comparisons.

Strategy #3: Convey the dynamic nature of data input streams and analytics outputs

Unlike publications, datasets that involve continuous measurement (e.g., in situ measurement of changes in soil properties) are dynamic. Attempts to generate insights from a limited scope of dynamic data can be misleading and have implications on the quality of outputs. When there are insufficient data points available to perform analytics and generate meaningful insights (e.g., large standard errors), this will be clearly stated as a limitation.

Strategy #4: Promote transparency in data processing and documentation

When working with raw data, researchers often make assumptions to fill possible gaps in the data and apply ad hoc rules to handle outliers. Although this is common practice, the steps in processing data must be clearly documented to ensure transparency and traceability. The Module will develop and promote best practice guides to support standard documentation of the workflows in data processing from the raw data. The raw data will remain openly accessible through the Data Exchange.

Strategy #5: Promote data citation and monitor the reuse and subsequent reanalysis of data

Each of the datasets provided through the official repository will be given a dedicated Digital Object Identifier (DOI). The pre-formatted citation data will be readily importable to reference management software. This will help researchers track who uses the data, and understand how the data are reused and reanalyzed, and monitor whether the data are properly interpreted and used in subsequent analyses. This will also provide a way to communicate with the data users if any correction or update is necessary.

3.1.1.4 System linkages

Consortium-wide coordination will be critical to achieve the overall Objectives and Targets in this Module 1, Organize. During the initial phase, important decisions will be made, especially on the review and selection of enabling environment elements and the common analytics environment. CGIAR-wide efforts on OA/OD compliance and scientific quality assurance will also need to continue. There will be largely three levels of inter-connected coordination mechanisms launched, each with detailed accountability framework. See Module 2, Convene, for more detailed plans to effectively implement System Linkages.

Linkage #1: Community of Practice on the OA/OD Compliance

Coordination on OA/OD compliance will be built using the momentum of the ongoing OA/OD initiative working with existing CGIAR data and knowledge management CoPs. Led by the Platform secretariat, this will extend to include representatives from all CRPs, setting shared objectives and milestones to achieve their full OA/OD compliance. Activities to support researchers in complying with the OA/OD policy will be also coordinated with the Platform's lead. Additionally, CGIAR's participation in global OA/OD collaborations and initiatives, such as GODAN (Global Open Data for Agriculture and Nutrition) and RDA (Research Data Alliance) will be coordinated through this linkage.

Linkage #2: Community of Practice on Data Analytics Users

Coordination on the technical aspects of applying data analytics will be implemented through a Community of Practice on Data Analytics, including CGIAR Centers and the Platform's academic partners (e.g., University of California – Davis, University of Florida, University of Nebraska, and

Wageningen University). The Platform's will support CRPs' scientific advances by providing guidance for generating data-driven insight through partnership for exchanging ideas, building capacity, and sharing research data across institutions. This partnership will also collaboratively explore and articulate ways to handle sensitive data and associated issues – such as legal, ethical, and privacy concerns.

Linkage #3: Big Data-driven Impact Assessment of CRPs

As the SDG targets are set for 2030, it is anticipated that a wide range of research activities across the CRPs will have the potential to contribute to monitoring the progress of SDG indicators, especially those aligned with CGIAR Goals as identified in the SRF. This necessitates CRP-wide technical coordination to ensure effective big data analytics approaches, and alignment around indicators. In close consultation with Agri-Food System Programs (AFS CRPs) and the Monitoring, Evaluation, and Learning CoP that includes all Centers and CRPs, the Platform will support the development of a series of case studies with products designed as relevant for each CRP's IDOs and Sub-IDOs, using multidisciplinary big datasets and ICT-assisted rapid data collection and analytics tools. This will facilitate the effective monitoring and progress of agricultural sustainability at scale. Case studies will also fully utilize information provided by the Global Integrating Programs; the Platform is intended to play an intermediary role between AFS CRPs and GIPs.

3.1.1.5 Climate Change

Linkage #1: Ensuring Standards Compliance on Climate Data

High-resolution climate data, both historic baseline and future projections, are already publicly available through the [CCAFS Climate Data Portal](#) and being used widely. To further facilitate the use of this resource, the Platform will provide guidance for enhanced discoverability and interoperability of the data (e.g., incorporating metadata standards such as CG Core) and explore options for easier access by machines through the implementation of Linked Open Data protocols and Application Programming Interfaces (APIs).

Linkage #2: Network of Micro-climate Sensors at Trial Sites of CGIAR and Partners

Monitoring of current micro-climate conditions (e.g., soil moisture and weather variables) can provide useful insights into climate variability and its changing spatio-temporal pattern – leading to a better understanding of climate change. There is unexploited potential to aggregate field-measured crop health and environmental data from across CGIAR trial sites, shared through open platforms in real-time, and linked with climate change analytics – for example, to evaluate performance of new technologies in different agroecological and climate-smart agriculture contexts. Coordinated by CIAT's AgTrials initiative in collaboration with connected sensors and weather station expertise in the Platform partnership, data streams from connected sensors will be publicly available through the CGIAR Data Exchange.

3.1.1.6 Capacity development

Big data analytics is currently not a mainstream research tool in CGIAR. As addressed in the objectives of this Module, opening CGIAR's research data will be the first step to take, providing researchers with key datasets, enabling tools and services, as well as a common analytical framework. These will all synergistically help bring big data analytics and data-driven insights to the core business of CGIAR and partners. Realizing the full potential of this synergy at large scale will require a consortium-wide

coordination to provide multiple levels of capacity development to the primary group of users across CGIAR and beyond. Through the convening opportunities addressed in Module 2 of this proposal, ample technical training opportunities will be provided via on-site and online courses. Shared resources such as Code Repository will also provide opportunities for researchers to learn from practical examples. As part of the training program, trainees will be encouraged to develop case studies where big data analytics are incorporated in their research, which can potentially feature in the Challenge Program described in Module 3.

In addition to the technical aspect of capacity building activities, full compliance for OA/OD in CGIAR and enforcing the policy across Centers and programs will require management-level transformation and commitment, necessitating high-level institutional capacity building. This may not be achieved quickly; a step-wise phased approach will be implemented across CGIAR that will further build on the momentum from the OA/OD Phase I initiative.

3.1.1.7 Intellectual asset and open access management

Overall, the CGIAR's OA/OD principle and policy will extend to all activities and products from this Module. For the key datasets to be offered through Objective 2, the Platform will serve as a clearinghouse for data access and management issues that may need to be addressed before making them publicly available through CGIAR Open Data Exchange. For other types of useful commercial data (e.g., commercial high-resolution remote sensing and weather forecast data), depending on the agreement with the provider, the Platform may elect to provide a limited scope of dataset as a sample in public, whereas costs will incur when its full access is necessary. In such case, the Platform will act as an intermediary between the data provider and CGIAR as a whole (similar to the mode of operation used by the Shared Services unit in the Consortium Office) to negotiate the terms while pursuing the best interests of CGIAR. More details on this is discussed in Module 2, Convene.

Privacy issues supersede open access and privacy sensitive data will be anonymized and/or aggregated prior to publication.

3.1.1.8 Module management

Overall activities of the Module 1 will be managed by IFPRI, in close collaboration with the Data and Knowledge Management Unit of the Consortium Office for the OA/OD-related targets and outputs in Objective 1 and the Shared Services Unit of the Consortium Office for consortium-level agreements to provide datasets, tools, and services across researchers in CGIAR. Identification of key datasets, tools, and organizing capacity building activities will be coordinated in close collaboration with external expertise especially in academia, including University of California – Davis (UC-Davis) and Columbia University's Center for International Earth Science Information Network (CIESIN).

3.1.2 Module Budget Narrative

3.1.2.1 General Information

Platform Lead Center's Name: CIAT

Module title: ORGANIZE

Center Location of Platform Leader:

3.1.2.2 Summary

Big Data Coordination Platform: Full Proposal 2017-2022

Funding Needed	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2	4,336,320.5	3,172,574.31	2,261,673.73	1,159,962.08	1,125,489.81	1,192,411.46	13,248,431.9
W3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	4,336,320.5	3,172,574.31	2,261,673.73	1,159,962.08	1,125,489.81	1,192,411.46	13,248,431.9

Funding Secured	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2 (Assumed Secured)	424,646	0	0	0	0	0	424,646
W3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	424,646	0	0	0	0	0	424,646

Funding Gap	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2 (Req	-3,911,674.5	-3,172,574.31	-2,261,673.73	-1,159,962.08	-1,125,489.81	-1,192,411.46	-12,823,785.9
W3 (Require	0	0	0	0	0	0	0
Bilateral (Fur	0	0	0	0	0	0	0
Other Source	0	0	0	0	0	0	0
	-3,911,674.5	-3,172,574.31	-2,261,673.73	-1,159,962.08	-1,125,489.81	-1,192,411.46	-12,823,785.9

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
Personnel	367,522.8	382,223.71	397,512.66	413,413.17	429,949.69	447,147.68	2,437,769.7
Travel	110,000	86,000	36,000	20,000	16,000	16,000	284,000
Capital Equipm	0	0	0	0	0	0	0
Other Supplies	612,419.94	481,636.81	378,291.82	220,966.82	178,966.82	218,966.82	2,091,249.01
CGIAR collabor.	2,077,500	1,548,250	926,250	106,250	106,250	106,250	4,870,750
Non CGIAR Coll	945,000	490,000	370,000	280,000	280,000	280,000	2,645,000
Indirect Cost	223,877.77	184,463.79	153,619.26	119,332.1	114,323.31	124,046.96	919,663.18
	4,336,320.51	3,172,574.31	2,261,673.74	1,159,962.09	1,125,489.82	1,192,411.46	13,248,431.89

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
CIAT	174,825	120,712.5	73,537.5	18,037.5	18,037.5	18,037.5	423,187.5
IFPRI	2,241,495.5	1,612,361.81	1,328,136.23	1,051,924.58	1,017,452.31	1,084,373.96	8,335,744.4
AfricaRice	125,000	92,500	50,000	0	0	0	267,500
Biodiversity Interna	340,000	289,500	220,000	90,000	90,000	90,000	1,119,500
CIFOR	125,000	92,500	50,000	0	0	0	267,500
ICARDA	125,000	92,500	50,000	0	0	0	267,500
ICRISAT	125,000	92,500	50,000	0	0	0	267,500
IITA	125,000	92,500	50,000	0	0	0	267,500
ILRI	205,000	132,500	90,000	0	0	0	427,500
CIMMYT	125,000	92,500	50,000	0	0	0	267,500
CIP	125,000	92,500	50,000	0	0	0	267,500
IRRI	125,000	92,500	50,000	0	0	0	267,500
IWMI	125,000	92,500	50,000	0	0	0	267,500
ICRAF	125,000	92,500	50,000	0	0	0	267,500
WorldFish	125,000	92,500	50,000	0	0	0	267,500
	4,336,320.5	3,172,574.31	2,261,673.73	1,159,962.08	1,125,489.81	1,192,411.46	13,248,431.9

Explanations of these costs in relation to the planned 2022 outcomes

Module 1 aims for the following high-level outcomes:

- 1) CGIAR Centers and programs further their public goods mandate through OA/OD compliance: A large portion of the Module 1 budget (about 50%) is dedicated to address needs articulated by Centers during the OA/OD Phase I, in support of consistent institutional implementation of the OADM Policy and monitoring and evaluation of progress. This budget will be frontloaded in Years 1-3 to sustain the momentum of Phase I and ensure the availability of open data sets to harness to big data capabilities. Sustained efforts are expected to continue as a core mandate of each Center and Program and through the culture change brokered by activities in Module 1 to foster data exchange and sharing.
- 2) Meaningful integration of CGIAR data sets achieved through semantic data management standards: Integrating agricultural data across disciplines, geographies, scales etc. in meaningful ways is essential to enhance innovation and impact. Semantic data management and interoperability enables information and data linking with unambiguous, shared meaning ascribed to the data, which in turn allows machine-computable logic, inferencing, knowledge discovery, and data federation among information systems. Activities in Module 1 that contribute to this outcome account for about 20% of the total budget.
- 3) Researchers are able to discover CGIAR data and leverage it through tools and services: CGIAR data, when made available, is typically uploaded to institutional repositories, many of which

still represent silos that are not easily accessible. There is no way of finding or connecting related types of data across these Platforms (e.g., rice data, publications, extension fact sheets for similar geographies or varieties from CIAT, IRRI, and AfricaRice). Activities accounting for about 20% of the Module 1 budget will facilitate CGIAR-wide discoverability of data, no matter where these data are archived, via an infrastructure provisionally envisioned as distributed cloud storage and services. Data discovery and reuse will enable review, assessment, and refinement of existing analytic tools and services, and development of new ones, further enriching the data, analytics, and innovation landscape.

- 4) Scientists work in a common data analytics environment: A community “one-stop shop” providing robust, proven analytics to seamlessly leverage well-described, quality controlled data sets is essential to fully empower researchers in the agricultural domain to develop innovative solutions to complex problems. Activities accounting for about 10% of Module 1 will contribute to such a common analytics environment; coupled with advocacy, training, tool development and promotional activities in the other Modules to truly bring big data to the community and the community to big data.

3.1.2.3 Additional explanations for certain accounting categories

Benefits:

A total of \$2,437,770 is budgeted for salaries and benefits for 4 employees (a Contracts and Grants Manager, 2 Senior Researchers, and a Research Assistant). Of this total cost, \$1,538,025 is budgeted for salaries and \$899,745 for Fringe Benefits.

IFPRI benefits are charged at a rate of 58.5% over the cost of salaries. The following is the breakdown of the fringe costs:

00.86% = Home Leave
 01.18% = Dependent Education/Travel
 13.69% = Disability/Life/Medical Insurance/Employee Assistance Plan
 00.17% = Workman’s Compensation
 00.26% = DC Unemployment Tax
 19.46% = Pension costs/Post Retirement Benefits
 20.93% = Leave
 00.99% = Relocation Expenses
 00.96% = Others

 58.50% = Total

Other supplies and services:

Of the total \$1,948,857 budgeted for Other Supplies and Services, the following is budgeted for:

- \$700,000 for meetings and conferences, including interactive regional OA/OD workshops for Center/program-specific technical discussions on OA/OD implementation and challenges, CGIAR-wide open access working group workshops to ensure consortium-wide OA/OD compliance on knowledge products, and agri-semantics meetings to advance semantic interoperability across CGIAR and collaborators.
- \$8,000 for general facilities expenses (minor software/hardware purchases, other supplies beyond the normal support provided by IFPRI).

- \$283,000 for advocacy and monitoring materials (Altmetrics, GODAN, ORCID, publishing open data), communications (training material), computer/peripherals/software for supporting the Module activities.
- \$15,000 for IFPRI communications cost for designing and printing publication materials (IFPRI Communications Division work specific to the program)
- \$942,857 for IFPRI Service Center Charges. This covers charges related to the infrastructure and support units responsible for providing the resources necessary to conduct the lead Center’s research agenda specific to Module 1. The cost of the services is allocated to benefiting projects based on utilization of these services by research staff. Utilization is measured by the number of direct labor hours incurred by research staff while providing direct research support under each project. Service centers are comprised of computer, library, facility, and research support services. The Computer Service Center supports research professionals by providing and maintaining computers necessary for the provision of research services. The Library Service Center provides reference resources for project support. The Facilities Service Center is responsible for maintaining the facility and includes the cost of rent, equipment, and facility support. Research support represents functions that provide support to the overall research capacity of the institute. The benefit of these services is shared across all research projects and is charged out at fixed hourly rates when research is performed by staff at IFPRI’s institutional offices.

3.1.2.4 Other Sources of Funding for this Project

In case only partial project funding is available, activities will be prioritized to achieve maximum compliance on open data across CGIAR, recognizing that unrestricted access to consistently and well-annotated, high quality data is the foundation this Platform is built on. If necessary, some envisioned meetings might be replaced with virtual meetings to reduce meeting and travel budgets. Some staff time may also be replaced with more in-kind contributions (see below). During the implementation of OA/OD Phase 1, for example, Centers were overall very supportive in trying to work towards a clear common goal despite OA/OD amounting to an unfunded mandate in the absence of direct financial support. Centers undoubtedly need financial assistance to build on the momentum of Phase 1 and deliver strong progress towards addressing gaps identified during Phase 1. However, following initial support to jump-start a strong sharing culture and outcomes, Center expenses for OA/OD can be potentially budgeted within CRPs and bilateral projects.

As data analytics on open agricultural data become mainstreamed and established as a key research method, we expect there will be many collaborative, in-kind contribution efforts identified across CGIAR and beyond it. In Year 1 (2017) for example, IFPRI’s Virtual Information Project will provide financial support and in-kind contributions to activities dealing with agricultural technology-related information and knowledge products (e.g., ontologies on agricultural technology, agricultural investments, geospatial datasets for technology targeting and adoption monitoring). Various OA/OD activities of Module 1 will further synergize with ongoing activities under the OA/OD initiative and CGIAR Center and CRP strategic efforts to comply with the Open Access policy as well as similar donor policies (BMGF policy, USAID policy and others). Through the OA/OD initiative, for instance, centers are developing implementation plans based on a common template to build or augment human and infrastructural capacity via overhead costs. Building OA/OD budgeting into bilateral project costs will

be especially crucial for securing funding in the future as donors increasingly mandate grantee adherence to open access and open data.

3.1.2.5 Budgeted Costs for certain Key Activities

Budgeted Costs for certain Key Activities

	Estimate annual average cost (USD)	Please describe main key activities for the applicable categories below, as described in the guidance for full proposal
Gender	100,000	Datasets made open access through Module 1 activities will be gender-disaggregated wherever applicable; this will be achieved in close engagement with CGIAR Community of Practice on Gender. Data standards developed in this Module will also include appropriate metadata for representing the gender dimension. Outcomes from Module 1 will help to democratize access to the information across gender. Overall, we estimate about 5% of total budget of this Module will be used to ensure this.
Youth	100,000	We anticipate that increased access to agricultural data and ICT-based applications will help retain and attract more rural youth to agriculture. Young men and women will be considered a major target user group for the annual data/knowledge consultations and design of communication materials. Overall, about 5% of the total budget of this Module will be allocated towards youth-targeted efforts.
Capacity development	654,000	Centers will receive dedicated funding for developing capacity to implement OA/OD and ensure science and data quality. It is anticipated that workshops will be also focus on or include capacity building as a core theme.
Impact assessment	160,000	Consultations with the MEL CoP and data/knowledge CoPs will take place to assess the overall impact of OA/OD and define the scope for identifying opportunities and monitoring progress towards impact. Start-up funds for Centers to implement Altmetrics and ORCID will help in monitoring.
Intellectual asset management	63,000	Staff time of 0.66 FTE/year is budgeted to handle issues related to legal aspects of OA/OD and IA management, legal counsel, and contract/grant management. Some in-kind support on legal aspects of OA/OD and IA management will also be provided in-house for Center-specific issues.
Open access and data management	2,200,000	The entirety of Module 1 is aimed at addressing open access and data management across CGIAR.
Communication	20,000	Development of advocacy and training materials, support for open agricultural data publications, and design/printing of communication materials are included.

3.1.2.6 Other

OA/OD is a relatively new challenge in the agricultural domain, but there are successful models to learn from in other disciplines such as the biomedical, genetics, genomics, and related molecular fields. For instance, [NCBI](#), conceived with a USD 4 million allocation from the US government in 1986 enables anyone with internet access to freely mine a vast collection of data and analytical tools, providing the sort of faceted search results and integration we envision for the agricultural domain. The discovery and reuse of data and publications via NCBI databases have resulted in greatly increased collaboration, and major advances in medical knowledge and impact—so much so that current annual allocation to NCBI is estimated to be over USD 200 million. The ecological informatics field has also seen OA/OD efforts such as the [Knowledge Network for Biocomplexity](#) (KNB), an international repository to enable the consistent annotation, sharing, discovery, and access of ecological data. Well-annotated and accessible datasets such as those in KNB allow researchers to find, integrate, and analyze diverse data to better understand and address complex challenges. The [Data Observation Network for Earth](#) (DataOne) is an initiative launched in 2009 with a USD 21 million award from the National Science Foundation (NSF) to provide interoperability and make content discoverable across a distributed system of repositories (including KNB). DataOne received another USD 14 million grant from NSF in 2014. Clearly the technology and infrastructure exist to accomplish easy discovery, sharing and reuse, integration, and analysis of scientific outputs, and other scientific domains have advanced from utilizing them fully. It is past time the agricultural domain harnessed these capabilities to accelerate innovation in addressing serious and complex global challenges for enhanced impact on

the ground. Given the budgets for similar efforts in other domains, a start-up budget averaging \$4-5 million over six years for this big data Platform represents a compelling value proposition to do so.

3.1.3 Module Uplift Budget

Uplift Budget

2022 ADDITIONAL OUTPUTS/SERVICE DESCRIPTION	AMOUNT NEEDED (\$)	W1 + W2 (%)	W3 (%)	BILATERAL (%)	OTHER (%)
1. REAL-TIME FARM MANAGEMENT DECISION SUPPORT SYSTEM WITH VOICE-RECOGNITION AND NATURAL LANGUAGE PROCESSING ENGINES, IN COLLABORATION WITH PRIVATE SECTOR Smallholder farmers need a basket of options that they can quickly sift through and implement. As a key determinant of productivity and profitability, weather is becoming increasingly unpredictable. Many of these farmers tend to be semi-literate or illiterate, and unable to interact effectively with smartphones. This project will leverage private-sector partnerships with advanced natural language processing, voice recognition, and artificial intelligence technologies, as well as advanced weather forecasting and remote sensing-derived products. It will provide useful, locally-relevant information to a smallholder farmer in response to her voice query, in the national language and as a voice recording. CGIAR's collective expertise on agronomy, socio-economy, and accumulated knowledge on farming will synthesize best farming practices for local farming conditions and will be used to train IBM's Watson artificial intelligence engine. Watson will be trained to understand the complexity of low-input agriculture in a holistic way and help provide context of local conditions based on the location of farmer. It will also mine data from other sources, involving a number of data sources that interact via a decision support application, with translation into and out of natural language. For example, if a question "When should I plant my rice?" is received through a phone (or app), it is translated and contextualized based on the caller's location. The agro-climatic conditions and weather forecast data will be pulled from weather database providers. Information on rice planting requirements will be pulled from an agronomic-knowledge database, soil conditions will be provided by remote sensing estimates, and market data retrieved from an agricultural market database. These data will be synthesized in an artificial intelligence-based decision support system to simulate multiple planting date scenarios and estimate the optimum date. The answer will be something like, "Rain forecast not sufficient for transplanting within next 2 weeks, recommendation is to direct-seed at 60	6,000,000	100	0	0	0

kg/ha in 2 weeks, with a basal application of 80 kg/ha urea, with a short-season variety.”					
2. SUPPORT FOR UTILIZING CLOUD-BASED COMPUTATIONAL INFRASTRUCTURE	1,080,000	100	0	0	0
There are many merits of utilizing cloud-based hosting and computation services for CGIAR scientists, especially those located in less-reliable on-site infrastructure. Cloud-based infrastructure provides on-demand scalable and disposable computational processing power for scientists. However, its pay-as-you-go pricing model has been a major block for its wider utilization. That is, until the task is complete, it is difficult for users to budget the cost upfront. To relieve this budget uncertainty and financial risk, this activity will provide a certain level of financial support to the scientists (up to 12K per center per year; flexibly manageable depending on the types of demand). Leveraging for the volume, the Platform will negotiate the best pricing model with private-sector partners and the industry-leader of cloud-based computing environment. As the cost of using cloud anticipated to drop over time, more scientists are expected to benefit from this support toward Years 4-6.					
3. SUPPORT FOR DATA QUALITY AND CURATION, DIVIDED AMONG 15 CENTERS FOR YEARS 1-3	3,000,000	100	0	0	0
For stakeholders in the agricultural research for development sector to share and consume data that is open, its quality needs to be assured. For data consumers in particular to reuse available data sets and keep the Big Data platform valuable and relevant, trust in the data will depend on factors that include: knowledge of the data source and its reliability, the relevance or currency of the data, and its quality and/or knowledge of the uncertainty that might be associated with it. CGIAR Centers and CRPs currently address such trust and quality issues in ad-hoc ways, depending on factors such as the resources available, and the unit to which data managers and curators belong. For instance, in some Centers and CRPs, data management is a part of, or closely associated with statistics/biometrics units or with a specialized data unit. As such, data sets to be made open typically undergo basic statistical analyses and/or simple mapping to assess the validity of numerical and geospatial values. CGIAR research focuses on different commodities, disciplinary domains, geographies and scales, resulting in very different types of data. Data types therefore range from agronomic trial data collected by field technicians in a variety of ways and formats, through input and output market information and socioeconomic data on adoption and drivers and conditions to enable it, to weather data and high-throughput sequencing and phenotypic information and satellite images. These datasets cannot all be treated in the same manner even if most data regardless of type are uploaded to one Center data repository. Their curation and quality control needs differ significantly—necessitating some level of statistical and/or subject expertise and customized approaches depending on the data type. Depending					

on receipt of the uplift support requested here, the primary responsibility for quality control and curation of data will continue to rest with Centers as the administrators and arbiters of CGIAR data repositories while the Big Data platform will undertake to harmonize approaches around quality and curation through efforts that include: – Guidance and training for CGIAR researchers and partners, for whom quality is a concern as their data increasingly becomes widely accessible and available for reuse, and for data managers/curators. – Recommendations on digital data collection tools, and negotiations and purchase where appropriate through a shared services model. – Operationalization of standards to enhance the quality and annotation of CGIAR data from collection onward (e.g., standard vocabularies, ontologies, and CG Core metadata schema). – Development and/or sharing of online tools, forms and templates for robust data collection and management (e.g., Breeding Management System; in-development Agronomy Management System and field book). – Development and sharing of workflows, tools, recommendations, and customizable Standard Operating Procedures for improved and consistent data management and quality.					
4. SUPPORT FOR ROBUST DATA SCIENCE WORKING ACROSS 15 CENTERS FOR YEARS 4-6	2,000,000	100	0	0	0
Recognizing that the Big Data platform has an ambitious vision in terms of bringing together harmonized data from different disciplines with state of the art visualization and analytical capabilities, and drawing lessons from public (NCBI) and private sector (IBM, Syngenta) efforts in this area, support is requested for appropriately skilled data scientists with subject and IT expertise to work across all 15 centers in Years 4-6.					
5. SUPPORT FOR CREATION AND MAINTENANCE OF ONTOLOGIES TO ENABLE DATA INTEROPERABILITY AS SHARED SERVICE, TO BE BASED AT BIOVERSITY/IFPRI FOR YEARS 1-6	1,000,000	100	0	0	0
CGIAR has made strong progress towards standards compliance to enable interoperability across repositories and related data sets. For instance, testing of the CG Core Metadata Schema by Centers/CRPs is underway; full implementation following any refinements necessary will contribute to discoverability and data integration. However, substantial work will be required to ensure that CGIAR platforms move beyond the metadata harvesting capability enabled by the Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) which is generally the norm, to data harvesting and the semantic interoperability offered by ontology-based and Linked Open Data (LOD) approaches. Uniform adherence to the improvement and consistent adoption of “agrisemantics” is required for improved data quality, annotation, resource discovery, and					

interoperability across CGIAR repositories, databases, and platforms. Agrisemantics efforts at CGIAR will focus on improving and/or building existing and envisioned ontologies (e.g. Crop Ontology, Livestock Ontology, Gene Ontology, Plant Ontology, Geo OWL, agricultural technology ontology, value chain ontology, fishery ontology). Funding is requested for Years 1-6 to enable this work as a shared service across CGIAR, to be undertaken by strengthened ontology teams already based at Bioversity and IFPRI.					
6. SUPPORT FOR HARMONIZATION AND ANALYTICS TO LEVERAGE HIGH-THROUGHPUT PHENOTYPING DATA, IN CLOSE COLLABORATION WITH GENETIC GAINS PLATFORM The Big Data platform seeks to accelerate innovation and impact by sharing data that is consistently annotated and harmonized for relatively seamless integration with related data and with analytic and visualization tools and services. Accomplishing this entails working across Centers, CRPs, and the other two cross-cutting platforms, and acknowledging the diverse curation needs of different data types. It requires a highly collaborative approach that includes agronomists, breeders, geneticists, computational scientists, and others working to ground-truth and adopt a coordinated approach to standards, protocols and best practices for collecting, managing, storing, sharing, and interlinking genotypic and phenotypic data from high-throughput systems and traditional systems. While some of the training and curation activities associated with genotypic and high-throughput phenotypic data are the domain of the Genetic Gains and Genebanks platforms, it is anticipated that the Big Data platform will need to work closely with these to collaboratively develop and ensure adoption of harmonized standards and approaches capable of consistently describing large volumes of non-traditional file types (e.g., the “heavy” image files arising from high-throughput phenotyping and remote sensing). A modest budget is requested to complement the efforts of the other platforms, recognizing that only then will these data be easily integratable with other related data (agronomic, socioeconomic etc.) and computational and decision support tools and applications.	600,000	100	0	0	0
7. SUBSCRIPTION OF ONLINE DRONE IMAGERY PROCESSING AND COLLABORATION PLATFORM AS A SHARED SERVICE FOR ALL CGIAR CENTERS Utility of drone-captured very-high-resolution of real-time remote sensing imagery in agriculture is expected to rapidly rise in the near future, from the use in high-throughput phenotyping, monitoring of crop performance (e.g., plant stress detection, yield mapping), enhanced mapping of ground features (e.g., farm plot boundary, farm animals), farmers’	100,000	100	0	0	0

adoption of technologies (e.g., use of machinery, field preparation practices, water managements), and precision agriculture technologies (e.g., yield mapping, soil nutrition management). During the technical consultations, it was identified that the main bottleneck of incorporating drones in the research is the difficulty of processing the imagery (i.e., special software needs) and the archiving and sharing of imagery, due to the significantly large size and volume of the data. Instead of implementing a solution for each center, we propose to subscribe an online platform where the drone-acquired data are uploaded, processed, and openly shared across CGIAR for further utility.

MODULE 2: CONVENE

3.2.1 Module Narrative

3.2.1.1 Rationale, scope

CGIAR is largely invisible in the broader field of big data analytics and ICT development. This is mainly due to the fragmentation of activities, and the lack of a concrete strategy to engage with the data analytics community. A key component of the theory of change of the Platform centers around the establishment of new partnerships, which can only be achieved if CGIAR elevates its visibility and status in the global community on data science. This activity will focus on bringing the big data community to agriculture, and show that exciting science and business opportunities exist in rural development and agriculture in developing countries. Positioning CGIAR as the go-to institution for data analytics in agricultural development will only strengthen the delivery of the ambitious goals in the SRF.

All 15 CGIAR Centers and all 12 CRP proposals contain elements of data analytics, applied to diverse fields from forestry to agriculture, from crop improvement to natural resource management, and from biophysical to social sciences. They manage data types that include economic surveys, geospatial datasets, genomic and phenomic data through to field trial and commercial farmer data. Some datasets are primary; others are secondary, originating from partner organizations. This great diversity of fields, data types, sources – and even more importantly, the analytical techniques then applied to them – represent a world of opportunities for collaborative, ground-breaking science. Yet many CGIAR scientists are unaware of efforts underway in other Centers, or CRPs they are not actively involved in. A key need for the cross-cutting Platform is therefore to convene scientists from across Centers, CRPs and key external partners and create the spaces that will stimulate greater collaboration, be it through data management and data sharing through to collaborative science.

This Module focuses on enabling those interactions to take place between different levels and programs within CGIAR, and between CGIAR and new partners in the big data field. A number of strategies are proposed to address this:

Big Data in Agriculture Convention: Central to enabling interaction is the concept of an annual Big Data Convention, bringing representatives from Centers and CRPs and partners to share information, develop joint initiatives and partner with key external actors in the big data space. The convention will differ from a standard conference approach, and will focus on hands-on interaction, partnership building, brainstorming and innovative sessions and approaches such as Hackathons and competitions. It will also be the basis for planning and sharing experiences of Inspire projects (Module 3) and the Steering Committee will meet around this event. Keynotes will come from high-profile researchers from both inside and outside CGIAR, and the aim is for the event to become a mainstay for any institution interested in developing big data applications in the context of agriculture in the developing world.

Communities of practice (CoP) and Working Groups (WGs): Establish, or work with existing CoPs around key topics that multiple Centers and CRPs are working on with sub-optimal interaction. Each CoP will be facilitated by a natural leader on the topic from within CGIAR, but will also reach out to external partners to participate and leverage the knowledge on Big data management and analytics available with partner institutions to enhance the potential for CGIAR impact.

Shared services and big data analytics infrastructure: CGIAR requires analytical tools and big data infrastructure in order to deliver on the promise of big data. Some of the private sector partners involved in this Platform can provide high-end services that would benefit CGIAR. Collective negotiation to access these services, such as IBM's data analytics capacity, or Amazon's cloud infrastructure, would greatly enhance CGIAR's capacity on big data and contribute to economies of scale, greater value for money and an improved enabling environment for Centers and CRPs to deliver on ICT or big data related outcomes and impacts.

Capacity Building: Capacity-building workshops will be held around the annual Convention on key big data and ICT topics in the field of agricultural development. These workshops will bring together CGIAR participants and partner institutions, including national partners, local universities and innovation hubs. Additionally, quarterly webinars will be organized for in-depth science presentation around specific topics (e.g. big data phenomics, crowdsourcing, use of drones in agricultural development etc.).

Collaboration Spaces: In addition to the physical convening of WGs and CoPs, virtual collaboration spaces, and a dedicated blog will be organized and provided for sharing information about ongoing initiatives. Connectivity will also be assessed to ensure that the communities are able to connect, share and collaborate easily. An annual assessment of tools to enable CoP function will be performed under this Module.

Through these different strategies, this Module will increase the capacity of CGIAR to engage on big data in agriculture, and broaden CGIAR's reach with non-traditional partnerships in this domain.

3.2.1.2 Objectives and targets

OBJECTIVE #1. Establish a CGIAR-led Big Data in Agriculture Convention and engage CGIAR more broadly in the big data community

This objective focuses on promoting greater sharing of experiences, opportunities, and cutting-edge science around big data. This includes taking CGIAR experiences into the wider world of big data science and implementation, as well as bringing new partnerships to CGIAR to improve big data capacity. The primary activity under this objective is the organization of a Big Data in Agriculture Convention, which will be an ambitious, high-profile event designed to be the go-to-meeting for any individual or institution interested in engaging on big data in agricultural development. The meeting will be designed to be highly dynamic, and focused on partnership building and brainstorming. Innovative hackathons and killer-app competitions will also be organized on the sidelines, and sessions will be organized to also develop Inspire project ideas (see Module 3 for more details).

Secondary activities include CGIAR presence in key Data Science [gatherings](#) and key focused disciplinary scientific gatherings where appropriate, such as the triannual conference of the [IAAE](#) or the [ACS](#) meetings.

MODULE 2. CONVENE	
OBJECTIVE 1. Establish a CGIAR-led big data in agriculture convention and engage and promote CGIAR more broadly in the big data community (40%)	
Target 1.1. Bring new partners to the table of CGIAR around big data analytics and promote CGIAR big data science	
Output 1.1.1.	Organization of sessions on big data in agriculture in broader big data scientific conferences and events.
Output 1.1.2.	Innovative communication efforts including the use of prizes for celebrating success stories on big data in agriculture, and participation of CGIAR teams in big data awards, prizes and initiatives.
Output 1.1.3	Establishment of strategic partnerships with African Innovation Hubs to promote CGIAR partnerships with entrepreneurial African companies and organizations.
Output 1.1.4	Aggressive communication of CGIAR success stories using big data in agriculture, including photo stories, videos and social media, and through CGIAR Big Data and ICT blog.
Target 1.2. Hold an annual Big Data in Agriculture Convention and establish as a go-to-meeting to engage in this topic	
Output 1.2.1	Annual Big Data in Agriculture Convention, involving participants from across CGIAR, Platform partners and other institutions engaged in big data and agricultural development.
Output 1.2.2	Organization of “hackathons” to stimulate new analytical approaches on CGIAR datasets. Broad dissemination of new analytical approaches as global public goods.
Output 1.2.3	Two training programs annually for CGIAR and partners on priority topics, including data management, cutting edge data analytics, ICT development etc.

OBJECTIVE #2. Establish communities of practice and working groups around key topics

Seven topics are identified where CGIAR-wide knowledge management and collective action is deemed a priority. Greater coordination across CGIAR, and the injection of new ideas and ways of doing business by partner organizations should increase CGIAR’s capacity to work on these topics more quickly, more effectively and at greater scale. CoPs create critical mass not achievable at individual Center level. The specific terms of reference for each CoP will be established by the Secretariat, but tasks would include:

- Needs assessments of CRPs and Centers around specific CoP or WG topic;
- Identification of best practices across CGIAR, and promising experiences from outside CGIAR, and sharing of knowledge across Centers/CRPs;
- Sharing of models, scripts, tools, platforms to reduce duplicity and maximize efficiency; and
- Establishment of new collaborative research efforts and proposal development for bilateral funding to advance the science.

An institution is assigned leadership, but their role is only to facilitate the community of practice which should be open to all CGIAR Centers/CRPs, and partners on this proposal. External expertise on these topics is especially crucial to identify new ways of doing business for CGIAR. When appropriate, the Platform will build off existing initiatives within CGIAR (e.g. the Data Management Task Force).

MODULE 2. CONVENE	
OBJECTIVE 2. Establish communities of practice (CoP), and working groups (WG) around key topics (60%)	
Target 1.3. Bring new partners to the table of CGIAR around big data analytics and promote CGIAR big data science	
Output 1.3.1.	CoP on Crop Modeling: Support the convening of CGIAR-internal Community of Practices on Crop Modeling, for its special focus on the development of practical demand-driven big data analytics use-cases. This will link with the ongoing effort from AgMIP. Led by IFPRI.
Output 1.3.2.	CoP on Spatial Analysis: Support the convening of CGIAR Consortium for Spatial Information (CSI), to share experiences working with spatial datasets, establish agreements for spatial data management standards and platforms, and enable a data sharing culture across CGIAR. Led by IFPRI.
Output 1.3.3	CoP on Data-driven Agronomy and ICTs: Share experiences across Centers on data-driven agronomy approaches and ICTs developed for farmers themselves, and establish collaborative research around the use of big data analytics to generate site- and climate- specific recommendation domains for agricultural development. Led by CIAT.
Output 1.3.4	CoP on economic survey data: This community of practice will share experiences and explore and develop novel opportunities for collecting, managing, storing, integrating (vertical and horizontal) and sharing high variety and high variability economic data and socio-economic survey data. The CoP will keep the minimum set of requirements that any data collection regarding human subjects and communities must abide by up to date. Led by CIMMYT.
Target 1.4. Establish working groups to agree CGIAR wide standards and best practices around key topics	
Output 1.4.1	WG on Data Management: Leveraging of a working group across CGIAR for defining data standards and inter-operability protocols, building off the Task Force on Data Management that currently operates. This working group will collaborate with other agencies and initiatives such as GODAN. Led by Medha Devare. (IFPRI/System Office)
Output 1.4.2	WG on Data Products: Identification of next generation gold standard big datasets that CGIAR could bring together, building on previous successes (e.g., WorldClim and SRTM) and providing supports for the convening. Lead TBC.
Output 1.4.3	WG on Infrastructure: Needs assessment across CGIAR for shared services to enable improved big data analytics (processing capacity, storage capacity, high-end analytics support etc.). Establishment and negotiation of agreements for access to big data infrastructure (storage, processing and data sharing amongst others). Led by Secretariat.

3.2.1.3 Science quality

This Module is not research *per se*, but rather the establishment of an enabling environment for excellence in big data science and analytics. The Communities of Practice and Working Groups are being assigned to established experts within CGIAR who have strong track records on the topic, and whose role will be to ensure that the Community of Practice discusses methods and approaches that are at the vanguard in terms of scientific quality. Hence, the CoPs and WGs themselves will have the explicit role of ensuring science quality on the selected topics across CGIAR.

The design of the Convene Module is to ensure that we develop a "learning organization" (Senge, 2006). A learning organization unifies around the themes of: systems thinking, personal mastery, mental models, shared vision, and team learning. In turn, scientific quality will be assured through these same themes:

- Systems thinking – the Big Data Platform will be treated as a bounded object within the scope of the broader mandate of CGIAR. By conceptualization both the elements of the Platform, we will also have clear demarcation for what it is not. This systems conceptualization will help minimize excessive scope while, at the same time, identify the primary channels serving to connect the Platform to the broader umbrella organization.
- Personal mastery – as mentioned, the CoPs and WGs are to be led by experts with established scientific track records and demonstrated ability for lateral thinking. Learning is a highly individual and incidental process, thus we look to the CoP and WG leaders to translate collective individual learning to organizational learning.
- Mental models – while systems thinking requires the development of a common understanding of the bounds of the Platform, the Platform recognizes that each of the participants brings a unique set of assumptions regarding not only what is Big Data, but how it applies in the context of agriculture and development. These assumptions represent values that may be shared, may be unlearned, or may be redoubled as the Platform interactions develop. The Convene Module will facilitate dialog on individual and collective mental models to arrive at a shared vision.
- Shared vision – Convene will promote a shared vision regarding the potential for both big data in agriculture, but also the Big Data Platform in CGIAR. This shared vision will provide a basis for communication and dialog with stakeholders both directly and indirectly involved in Platform activities.
- Team learning – the purpose of the CoPs and WGs, and the SC is to facilitate rapid, multi-directional learning. The leaders and team members are facilitators in the team learning process. We will evaluate learning dynamics and related behavioral change throughout the life of the project using a variation of the [structured 5Q approach](#) developed at CIAT for rapid evaluation of project impact.

Through approaching the Convene Module through an organizational theory perspective from the very beginning, we set the stage for continue improvement throughout the life of the Platform.

3.2.1.4 System linkages

All Centers and CRPs will be actively encouraged to participate in CoPs and WGs that relate to them, and will be invited likewise to participate in the annual Convention. Each CRP and Center will be asked to provide a representative through whom the Platform will interact. Two-way dialogue between the Platform and CRPs is essential – the Platform is designed to be a support mechanism to ensure compliance of OA/OD principles, and create an enabling environment for CRP research to use big data and analytics more effectively to solve CRP-specific problems and achieve targets. Under this Module, the CoPs and WGs are proposed as effective vehicles for these linkages to be made between the Platform and CRPs (and by default the Centers involved).

Some CoPs and WGs will have greater relevance for certain CRPs over others. For example, the socio-economic CoP will be especially important and relevant to collaborate with PIM, in addition to the socio-economic Flagships in AFS-CRPs. The Secretariat is charged with ensuring that the CoPs maintain appropriate inclusion of key CRP and Center representatives. The CoPs and WGs will also actively

involve non-CGIAR partners to leverage knowledge and expertise and create critical mass. Through exploration of new ways of doing business, CGIAR can become much more effective.

Shared services (e.g. cloud infrastructure, high-end processing capacity) will be made available to all elements of CGIAR system, including CRPs, Centers and Platforms. The WG will be charged to gauge demands across the system for such shared services, and identify appropriate partnerships to deliver on those needs. The WG will be responsible to manage the resources made available to each CRP and Center should demand outstrip supply.

No CoP is proposed for management of genetic data and information as this is covered by the Genetic Gains Platform proposal.

3.2.1.5 Climate Change

Climate related challenges and data needs occur in various elements of this Module. Crop modelling is a significant user of climate information, and the source for impact studies to help guide prioritization of climate adaptation. CCAFS Flagship 1 is to be closely involved in the Community of Practice on crop modelling, and Flagships 2 and 4 to be involved in the Community of Practice around data-driven agronomy as climate is a key variable in the generation of agro-advisories. Jointly, this Platform should contribute to improved climate change foresight. The WG on data products may identify priority climate related datasets for generating new data products, following on from very successful IPG data products such as WorldClim or CCAFS-Climate. Partners such as TAHMU bring significant expertise and meteorological infrastructure to the Platform.

3.2.1.6 Capacity development

This Module is almost entirely focused on capacity development. The elements of capacity building considered under this Module are:

- Increasing CGIAR capacity on data analytical approaches through targeted capacity building workshops from identified strategic partners
- Increasing CGIAR partner capacity to use big data analytics in agricultural development
- Increasing organizational awareness of agricultural challenges and opportunities for big data application in institutions with a track record in big data analytics

These will be achieved through both formal capacity building workshops on key topics, identified on an annual basis by the Secretariat, and approved by the Steering Committee, or through more informal processes such as through participation in the various WGs and CoPs.

3.2.1.7 Intellectual asset and open access management

The WG on Data Products will identify the next generation of International Public Good data products that CGIAR can and should be generating. New data products will be developed from start to finish abiding by IA principles, and be maintained in a regularly updated IP portfolio (which in the Platform's case will be lists of information products such as publications, databases, analysis tools (e.g. big data models), and web services (e.g. modelling platforms)). Where raw data underlying data products contain privacy and sensitive information, the data will be anonymized and/or aggregated in order to protect the privacy of individuals involved. Development of new data products must abide by ethics guidelines.

The Communities of practice that are part of the platform will promote guidelines for ethics in research, and ensure protection of the privacy of all subjects involved in the data.

3.2.1.8 Module management

This Module is to be led by the Secretariat, but a number of responsibilities are to be distributed across CGIAR Centers/CRPs and external partners. CoPs will be led by experts in that domain as identified in the activity tables. Each CoP leader will be assigned US\$100k to facilitate dialog, exchange and ensure delivery of the terms of reference. These funds may be used for specific meetings, synthesis activities, and development of joint publications or generation of new data-related products or models. The WG on Data Products (lead by CIAT) and the WG on Data Management (lead by IFPRI) will be assigned US\$200k each to ensure delivery to their respective terms of references. Additionally, up to US\$200k will be assigned for establishing shared service agreements for infrastructure and analytical support that all CRPs and Centers can tap into under the WG on Infrastructure. These agreements will be negotiated and managed through the Secretariat.

3.2.2 Module Budget Narrative

3.2.2.1 General Information

Platform Lead Center's Name: CIAT

Module title: CONVENE

Center Location of Platform Leader:

3.2.2.2 Summary

Funding Needed	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2	1,452,300	1,512,927	1,576,295.64	1,643,044.96	1,712,693.32	1,785,438.3	9,682,699.21
W3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	1,452,300	1,512,927	1,576,295.64	1,643,044.96	1,712,693.32	1,785,438.3	9,682,699.21

Funding Secured	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2 (Assumed Secured)	0	0	0	0	0	0	0
W3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	0	0	0	0	0	0	0

Funding Gap	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2 (Required	-1,452,300	-1,512,927	-1,576,295.64	-1,643,044.96	-1,712,693.32	-1,785,438.3	-9,682,699.21
W3 (Required from	0	0	0	0	0	0	0
Bilateral (Fundraisi	0	0	0	0	0	0	0
Other Sources (Fui	0	0	0	0	0	0	0
	-1,452,300	-1,512,927	-1,576,295.64	-1,643,044.96	-1,712,693.32	-1,785,438.3	-9,682,699.21

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
Personnel	410,000	421,200	432,774	444,736.98	457,104.59	469,893.21	2,635,708.78
Travel	60,000	63,000	66,150	69,457	72,931	76,575	408,113
Capital Equipment	50,000	52,500	55,000	58,000	61,000	64,000	340,500
Other Supplies and Se	510,000	534,000	559,250	585,768	613,560	642,744	3,445,322
CGIAR collaborations	300,000	315,000	330,750	347,288.5	364,652.63	382,886.16	2,040,577.28
Non CGIAR Collaborati	0	0	0	0	0	0	0
Indirect Cost	122,300	127,227	132,371.64	137,794.48	143,445.1	149,339.94	812,478.15
	1,452,300	1,512,927	1,576,295.64	1,643,044.96	1,712,693.32	1,785,438.31	9,682,699.21

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
CIAT	1,143,300	1,188,477	1,235,623.14	1,285,337.8	1,337,101.11	1,391,065.55	7,580,904.6
IFPRI	102,000	107,100	112,455	118,077.75	123,981.64	130,180.72	693,795.11
CIMMYT	210,000	220,500	231,525	243,102.3	255,257.1	268,020.9	1,428,405.3
	1,455,300	1,516,077	1,579,603.14	1,646,517.85	1,716,339.85	1,789,267.17	9,703,105.01

Module 2 establishes mechanisms for coordination, collaboration and joint-research across CGIAR and with its external platforms. The total budget over 6 years is US\$9.68m. This budget remains relatively stable from Year 1 through Year 6 with only inflationary adjustments to the costs.

US\$200k is budgeted to hold the Annual Big Data in Agriculture Convention. This includes budget for hosting, sponsoring travel for developing country partners, and the organization of the range of events that will surround it including Hackathons, competitions, Dragon's dens, plus Working Group and CoP meetings and the Steering Committee on the sidelines.

Each of the 5 CoPs (socio-economic data, spatial data, data driven agronomy, crop modelling) will be funded with US\$100k/yr. This is to be used at the discretion of the leader and based on the particular needs of the user community, but as a rule of thumb is being budgeted as 60% personnel time and 40% operations (travel and supplies and services).

The WG on Infrastructure is budgeted to US\$230k per year, and includes US\$50k for Capital purchases within the CGIAR (central servers, for example), and US\$180 for contracted shared services with external partners. Examples of such contracts may include cloud storage or data processing

infrastructure, or high end analytical support. A computer and data scientist is budgeted to oversee this work for the Platform.

The WG on Data Products is budgeted to US\$200k/yr, with 75% to cover salaries and 25% under supplies and services to hold workshops or purchase data services. The exact distribution of this budget will depend upon the data products that are prioritized by users.

The WG on Data Management consists of US\$50k/yr for national staff support plus US\$100k/yr for travel and workshops to enable the data management task force to engage on specific priority topics.

3.2.2.3 Additional explanations for certain accounting categories

Benefits: National Research Staff (NRS): Fringe benefits for national staff (costs for all benefits are added to the base salary to provide the total cost of the position) are comprised of legal benefits (local mandatory) and extralegal benefits (CIAT mandatory) and the provisions to cover local legal requirements such as: Pension - social security, training and development, occupational health, transportation costs and subsidies, work clothes and personnel protection requirement, and food subsidy.

International Research Staff (IRS): Fringe benefits for international staff (costs for all benefits are added to the base salary to provide the total cost of the position) are comprised of housing allowance, education allowance, car allowance, Cost Of Living Allowance (COLA), hardship, home leave tickets, insurance, retirement contribution, occupational health, training and development, repatriation and relocation provisions.

Other supplies and services: Discussed above.

3.2.2.4 Other Sources of Funding for this Project

Sponsors will be sought to share the costs of the Convention and bring annual costs down. Local and international private sector companies will be invited to support specific events around the Convention. Staff time of participants to the CoP are provided in kind by their institutions. Many of the cost items under this module are about internal organization of the CGIAR, and is therefore very difficult to raise bilateral to support.

3.2.2.5 Budgeted Costs for certain Key Activities

	Estimate annual average cost (USD)	Please describe main key activities for the applicable categories below, as described in the guidance for full proposal
Gender	50,000	Elements of the Convention will focus on gender related topics from data management through to big data use in practice. Female participation in the convention will also be incentivized to break typical male dominated technology barriers down.
Youth	50,000	Significant youth involvement from within the CGIAR and with partner organisations will be sought in both the Convention and in CoP to bring new ideas to the table.
Capacity development	800,000	The CoPs and WGs focus entirely on organizational Institutional Strengthening, and Organisational development, hence there is a heavy capacity building element in this module.
Impact assessment	0	Minimal impact assessment is envisioned under this module.
Intellectual asset management	50,000	Some intellectual asset management will be considered under the CoPs and WGs as relevant.
Open access and data management	800,000	A significant focus of both the Convention and the CoPs/WGs is around open access and data management.

Communication	200,000	A large component of the Convention is focused on innovative communication around the topic of Big Data and Agricultural Development.
----------------------	---------	---

3.2.2.6 Other

N/A

3.2.3 Module Uplift Budget

2022 ADDITIONAL OUTPUTS/SERVICE DESCRIPTION	AMOUNT NEEDED (\$)	W1 + W2 (%)	W3 (%)	BILATERAL (%)	OTHER (%)
8. EXPANDED SECRETARIAT FUNCTIONALITY The base budget of the Secretariat is just US\$300k, which only covers the most basic of functions. Under this uplift item, an additional US\$200k per year is envisioned for the secretariat, which among other things will allow the secretariat to: • Convene an International Advisory Board to meet twice per year (one virtual, one face-to-face) to oversee progress • Recruit a higher caliber Platform Coordinator, and increase communications and program support • Engage further in publications, events and synthesis products • Increase bandwidth for partnership management, networking and fundraising, especially with respect to other global and regional open data initiatives such as open harvest, AgroKnow amongst others.	1,500,000	100	0	0	0
9. ENHANCED COMMUNITIES OF PRACTICES WITH BROADER TERMS OF REFERENCE Under the base budget scenario, a community of practice receives 100k, and is tasked with: • Needs assessments of CRPs and Centers around specific CoP or WG topic; • Identification of best practices across CGIAR, and promising experiences from outside CGIAR, and sharing of knowledge across Centers/CRPs; • Sharing of models, scripts, tools, platforms to reduce duplicity and maximize efficiency; and • Establishment of new collaborative research efforts and proposal development for bilateral funding to advance the science. Under this uplift scenario, each of the 4 CoP will expand this work to generate best practice guidelines, develop tools and solutions to emerging problems, and further position CGIAR as a leader on the topic through partnerships, capacity development and production of new tools and approaches. 100k additional funds are provided to enable this.	2,400,000	100	0	0	0
10. ADDITIONAL COMMUNITIES OF PRACTICE FOR ONTOLOGIES AND ETHICAL CONSIDERATIONS	2,400,000	100	0	0	0

Two additional Communities of Practice are proposed to deal with 2 emerging topics of significant importance. As per recommendation from the CO and ISPC on ethical issues, and building off of Annex 8, a Community of Practice that evaluates and monitors the multi-faceted challenges on data and ethics. An additional CoP is also proposed to develop Agricultural Ontologies, and work with FAO, Agroknow and other initiatives to further develop ontologies to permit interoperability between systems and models. This is a crucial precursor to effective data management, and streamlined use of data services to solve problems. Each of these additional CoPs are budgeted at US\$200k per year.

3.3 MODULE 3: INSPIRE

3.3.1 Module Narrative

3.3.1.1 *Rationale, scope*

Big data has the potential to drive agricultural development, yet there are limited concrete cases that show how data-driven approaches to development yield results in terms of poverty reduction, enhanced nutrition or environmental benefits. New approaches need to be identified that make use of available data-sources and new analytical tools, and levelled at classic agricultural development problems. Furthermore, systematic learning of success and failures needs to be documented, and synthesized in policy-relevant documents to drive future investments in big data related agricultural innovations.

Most scientific breakthroughs originate from an idea that requires pilot funding to demonstrate its feasibility. Under this Module, the Platform will provide small annual grants of US\$100k pilot funding for high-risk, innovative big data and ICT ideas that have the potential to lead to new scientific discoveries, or enhance the efficacy of development efforts. Pilot funding will be anchored in the CRP structure, ensuring leveraging of the funding and high likelihood of further development of the ideas under the CRP in subsequent years.

During the Annual Convention, collaborative brainstorming spaces will be designed to encourage Centers and CRPs to engage with external partners to develop novel ideas that use big data to solve agricultural development problems and challenges. Short pitches of these ideas will then be made to a multi-partner panel and the Steering Committee. High-risk ideas will be welcomed. The best ideas will be selected for funding. Grants will be managed with minimal administrative burden, following a similar philosophy to the Grand Challenges initiative pioneered by the Bill & Melinda Gates Foundation.

The topics for such Inspire projects will be tightly defined, and the projects must be aligned to a CRP, and include non-CGIAR partners. This ensures that the projects are not stand-alone, and have the possibility to leverage into significant impact if successful and scaled out by the CRP involved. The inclusion of non-CGIAR partners will also contribute to new collaborative efforts, and contribute to the ToC by bringing new ways of doing business to CGIAR and generate greater capacity on the application of big data to agriculture.

Topics for the Inspire projects will be identified on a yearly basis through consultation with CRPs to identify problems within their annual workplans where big data analytics could solve intractable problems, or vastly increase the efficiency in problem solving.

Specific topics where the Platform will focus on include:

- Crowdsourcing data from the field on a mass basis
- Drones and image analysis for high throughput phenotyping
- Analytics to understand complex crop-livestock-climate-soil-management interactions
- Spatio-temporal modelling in real time using interoperable data streams for food security and agricultural monitoring

The Steering Committee will make the final selection of topics that get prioritised, but examples could include:

- Novel means to evaluate CGIAR impact through crowdsourcing and remotely sensed data;
- Opportunities to close gender-gaps through big data analytics and ICTs targeted for women and youth;
- Closing yield gaps through improved big-data driven agro-advisories;
- Improving national policy development through big data analyses of trends and spatial patterns;
- Novel tool development for climate resilience planning using CGIAR data integrated with other data;
- Crowdsourcing of pest/disease dynamics, and development of dynamic pest risk mapping and advisories
- Development of tools that strengthen the transformative power of the platform to further the CRP objectives by harnessing the foresight and capabilities of Big Data;
- Lowering transaction costs of data gathering for research and policy making;
- Harnessing the information from remote sensing in novel ways including crowd sourced ground trothing; and
- Development of applications for the most simple of smartphones to bring the benefits of Big Data to resource poor farmers.

Pilot Inspire projects will be evaluated after one year, and the ideas which prove to be feasible and show greatest potential for scale out will be identified through a robust selection process. Funding will be increased to US\$250k per year for a two-year period for full Inspire projects to bring the idea to scale in close coordination with the CRP to which the project is aligned. These projects will generate new analytical approaches, scientific outputs and high profile examples that demonstrate potential of CGIAR to lead the field in big data and agricultural development. M+E and impact assessment will be used to document efficacy of the approaches developed and will be accompanied by strong communications, outreach and fundraising efforts. At the end of the project cycles, synthesis and lessons learnt will be written up and published, alongside policy briefs and best practice guidelines for big data in agriculture. Inspire project participants will also be connected to the broader Platform network through collaborative tools and thus, the network and its potential for generation of new research and new resources will grow. This will be measured as part of the network assessment and mapping process.

3.3.1.2 Objectives and targets

OBJECTIVE #1. Establish pilot Inspire projects around big data related innovations for agricultural development

Under this objective, pilot Inspire projects will be developed, selected, contracted, reported upon and lessons learned communicated. Each year, the Steering Committee will identify 1-3 priority topics for pilot Inspire projects and issue a call to CGIAR Platform participants (CRPs and CGIAR researchers) to develop ideas together with external big data partners. Ideas will be further developed through facilitated ideas labs (or sandboxes as referred to in the UK) around the annual Convention on Big Data and Agriculture. Proponents will then pitch the ideas to a selection committee consisting of the Steering Committee and selected topical experts. Criteria will be established to evaluate, and the best project ideas selected for US\$100k pilot funding to develop the idea further and test in a specific context for a one-year period. In 2017, a total of five Pilot Inspire projects are earmarked for funding, with a further three budgeted for 2018, and two for 2019 and 2020. No budget is earmarked for pilots in 2021 and 2022.

Target 1.2 under this objective focuses on systematic learning from the pilot Inspire projects, and communication and outreach of best practices around the topics prioritized, drawing on the pilot

activities and a broader range of literature, projects and research outside of CGIAR. The outputs focus on providing guidance for what big data approaches work best, where and how. Such outputs would be useful for defining future research and development priorities for big data analytics and data-driven agricultural development investments.

MODULE 3. INSPIRE	
OBJECTIVE 1. ESTABLISH PILOT INSPIRE PROJECTS AROUND BIG DATA RELATED INNOVATIONS FOR AGRICULTURAL DEVELOPMENT (30%)	
Target 1.1. Establish pilot Inspire projects around big data related innovations for agricultural development	
Output 1.1.1.	Annual identification of priority Inspire project topics based on emerging literature, development trends and emerging technological opportunities.
Output 1.1.2.	Pilot Inspire ideas workshops held around the Convention, and selection process established to evaluate proposals.
Output 1.1.3	Selection, contracting and implementation of 2-5 Pilot Inspire projects annually.
Output 1.1.4	New proven big data driven solutions to agricultural development piloted, and methodologies and approaches made available
Target 1.2. M+E of pilot Inspire big data innovations, outreach and dissemination of emerging lessons	
Output 1.2.1	Communication of Inspire project innovations, outputs and outcomes including photo stories, videos and social media, and through CGIAR Big Data and ICT blog.
Output 1.2.2	Synthesis of pilot Inspire project successes and failures, and policy documents and best-practice guidance on future big data opportunities for agricultural development (2019).

OBJECTIVE #2. SCALE SUCCESSFUL PILOTS of big data related innovations for agricultural development into fully-fledged Inspire projects in collaboration with CRPs

This objective focuses on scaling out successful pilots, and robustly testing innovations from Objective 1 across broader contexts. Each pilot Inspire project will be evaluated by the Secretariat and external evaluators to assess the potential of the innovations developed. These evaluations will be structured around criteria such as potential for impact, scientific novelty, and value for money, gender and youth benefits, scalability amongst others.

The top evaluated pilot Inspire projects will then be considered for scale-out funding, consisting of US\$250k per year over a two-year period. Strong support from the CRP that the project is aligned to will be ensured, and co-investment by that CRP and/or bilateral donors also pursued. The Secretariat will co-design the project with the proponents, and the innovation will be tested at scale in CGIAR Site Integration countries with target CGIAR beneficiaries. Each project will develop an impact pathway and align results to CGIAR SRF, thus contributing to IDO and sub-IDO indicators.

Scaled Inspire projects will be implemented over a two-year period, and will be expected to produce IPGs such as new methodologies and approaches, new datasets, and new business models, and produce scientific outputs such as journal articles, policy briefs and guidelines. One project will start in 2018, an additional project in 2019, two additional projects in 2020 and a final project in 2021. This staging will facilitate projects to learn from each other and build off previous lessons learned.

Projects will develop rigorous ex-post impact evaluation designs that will include robust M+E, baselines and end-line data collection to evaluate the potential of each innovation to generate impact. Systematic M+E and learning activities within each of the scaled Inspire projects will contribute to further synthesis in 2019-2022. Although impact assessment will be implemented throughout the projects' life, most of the resources for this activity will be focused in 2021 and 2022. At the end of the 6 years, this Module will have generated a series of pilots, projects, innovations and robust impact assessments that will be synthesized and communicated and disseminated broadly to donors, governments and development actors.

MODULE 3. INSPIRE	
OBJECTIVE 2. SCALE SUCCESSFUL PILOTS OF BIG DATA RELATED INNOVATIONS FOR AGRICULTURAL DEVELOPMENT INTO FULLY FLEDGED INSPIRE PROJECTS IN COLLABORATION WITH CRPS (70%)	
Target 2.1. Implement scaled Inspire projects around successfully piloted big data related innovations for agricultural development	
Output 2.1.1.	Evaluation of pilot Inspire projects following criteria established by the Steering Committee, and selection of projects for scaling.
Output 2.1.2.	Implementation of 2 year scaled Inspire projects; production of IPGs around datasets, methodologies, approaches.
Output 2.1.3	Scaled big data innovations in CGIAR target regions; contributions to CGIAR IDO targets
Target 2.2. M+E, impact assessment and synthesis of big data innovations in agricultural development	
Output 2.2.1	Impact assessment of scaled Inspire projects, including quantification of return on investment (ROI).
Output 2.2.2	Synthesis of Inspire project successes and failures, policy documents and best-practice guidance on big data opportunities for agricultural development, summary studies on return on investment of approaches developed.

3.3.1.3 Science quality

Pilot Inspire projects will be evaluated by the Steering Committee and other invited evaluators based on a number of science-related criteria including novelty, innovation, robustness and reproducibility. To streamline the evaluation process and ensure that small Inspire projects have the potential to result in fundamental change in our paradigms regarding the use of data and analytics in agriculture, we will develop a simple and consistent evaluation rubric. The rubric will frame measurement of each of the above categories to provide a basis for evaluating each project, but will also facilitate a "meta" understanding of the collective set of projects to identify missed opportunities and under-represented areas. The Inspire categories will be evaluated in the following manner:

- Novelty and Innovation will be evaluated in terms of relevance. Does the project address a previously overlooked need? Does the project show potential to benefit society? Do the presented innovations require big data and analytic approaches? Are the proposed innovations made possible through the resources of the Big Data Platform? By asking each of these questions in a structured manner, the strength of the linkages between the Inspire project, the Big Data Platform, and the SLOs can be clearly established.

- Robustness and Reproducibility will be evaluated in terms of validity. Does the project demonstrate internal validity by clearly demonstrating awareness of potential alternatives and their corresponding strengths and weaknesses? Does the project demonstrate external validity through the development of generalizable approaches? Is the workflow of the proposed project clearly documented and applicable beyond the stated problem? Again, a structured set of answers will serve to quantify and qualify the linkages of each Inspire project to the overarching goals of the Platform. Recipients will be required to demonstrate a path to scale for a project to be successful.

The Pilot Inspire projects coupled with the structured approach to evaluating scientific quality thus serves a dual purpose. First, the SC will be equipped to allocate funds to the projects that show the greatest promise not only to support the goals of the Platform and the SLOs, but also for those projects that will serve as beacons for future efforts. Equally important, the structured evaluation framework will support a characterization of the scientific space covered by the Inspire projects in a comparable manner. This coverage can then be cross-walked with any number of counterfactuals (the SLOs, the Platform themes, etc.), to determine whether there is excessive overlap in the project foci or to identify targets of opportunity that should be specifically pursued with appropriate subject matter experts.

3.3.1.4 System linkages

All Inspire projects (pilots and scale-out) will require alignment with a CRP to ensure that they are fully embedded in the larger CGIAR research portfolio. This will ensure that successful pilots can be scaled-out through the CRP. At minimum, a CRP should endorse proposals for pilot Inspire projects, and proponents should be able to articulate how it contributes to that particular CRP. Ideally the project will be embedded within a much larger CRP-funded initiative or project under a specific flagship, and the pilot involve an innovative piloting of a new method in a context where the CRP already has established ongoing research with delivery partners.

For Scaled Inspire projects, greater CRP involvement will be sought to ensure a) relevance, and b) greater leverage.

Pilot projects will not be held accountable for delivery with respect to SRF targets, but scaled projects will develop impact pathways that are mapped to SRF IDOs and sub-IDOs, hence contributing to system level outcomes and targets.

3.3.1.5 Climate Change

Areas where big data can contribute to innovation in climate mitigation or adaptation will be identified via multi-stakeholder consultations and other means, and will be considered by the Steering Committee for the Inspire project calls for innovations.

3.3.1.6 Capacity development

This Module contains less capacity building activities than other Modules, although capacity building elements can be expected under the Inspire projects. Projects selected for scaling will be requested to outline their capacity building activities.

Capacity building needs will also be considered in the synthesis of big data innovations for agricultural development and incorporated in lessons learned and guidance on future investment on big data in agriculture.

3.3.1.7 Intellectual asset and open access management

All agreements and contracts made with inspire projects will comply with the CGIAR IA Principles. The inspire projects will be asked to commit to sound management of intellectual assets and IP rights. All IA generated with CGIAR Big Data Platform funds, including data and software, will be open access. They will have open licenses that maximize global accessibility and reuse, and all data products will have to be machine readable, interoperable, and reusable. All the outputs of the inspire projects will adhere to these principles.

Risks exist with respect to [privacy and cyber-security](#) when data on individuals is being considered. The protection of individuals and their privacy supersedes open access policy. Any private data from individuals (including household information from farmers) needs to be aggregated and or anonymized before make it public. In some cases where it concerns sensitive data this may be extended to communities. The CGIAR Open Access Policy allows for some exceptions to Open Access principles, including restrictions for sensitive data (for example, private data) and confidentiality obligations. Some of the products produced inx the Platform may fall into those exceptions.

New and innovative approaches for classifying, hosting and sharing sensitive data in secure, legal and transparent ways such as the Datatags system (Sweeney et. al) proposed by [Datatags.org](#) project will be explored. Such a system will allow for the classification of the level of sensitive information in data and would generate trackable and machine readable tags with privacy information. This will be crucial for ensuring privacy in a big data environment that includes cloud computing and combining of different datasets.

The inspire projects will respect Farmer rights as such when using traditional knowledge, prior informed consent of farmers will be sought and all outputs referring to traditional knowledge will provide appropriate credit.

3.3.1.8 Module management

This Module is to be led by the Platform Secretariat, although participation is expected from a large range of contributors from across CGIAR and from partner institutions. Contracts for Inspire projects will be issued from CIAT. M+E and impact assessment plans and strategies will be developed jointly by the IFPRI and CIAT IA teams.

3.3.2 Module Budget Narrative

3.3.2.1 General Information

Platform Lead Center's Name: CIAT

Module title: INSPIRE

Center Location of Platform Leader:

3.3.2.2 Summary

Funding Needed	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2	612,720	670,095.9	1,017,294.24	1,307,446.44	1,089,338.37	538,356.48	5,235,251.43
W3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	612,720	670,095.9	1,017,294.24	1,307,446.44	1,089,338.37	538,356.48	5,235,251.43

Funding Secured	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2 (Assumed Secured)	0	0	0	0	0	0	0
W3	0	0	0	0	0	0	0
Bilateral	0	0	0	0	0	0	0
Other Sources	0	0	0	0	0	0	0
	0	0	0	0	0	0	0

Funding Gap	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
W1+W2 (Required from I	-612,720	-670,095.9	-1,017,294.24	-1,307,446.44	-1,089,338.37	-538,356.48	-5,235,251.43
W3 (Required from I	0	0	0	0	0	0	0
Bilateral (Fundraising)	0	0	0	0	0	0	0
Other Sources (Fund	0	0	0	0	0	0	0
	-612,720	-670,095.9	-1,017,294.24	-1,307,446.44	-1,089,338.37	-538,356.48	-5,235,251.43

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
Personnel	22,000	22,440	100,918.8	102,937.18	104,995.92	107,095.84	460,387.73
Travel	5,000	5,000	13,000	21,000	21,000	21,000	86,000
Capital Equipment	0	0	0	0	0	0	0
Other Supplies and Servi	25,000	26,250	102,562.5	103,942.5	105,390	106,910	470,055
CGIAR collaborations	250,000	275,000	350,000	475,000	375,000	125,000	1,850,000
Non CGIAR Collaborator	250,000	275,000	350,000	475,000	375,000	125,000	1,850,000
Indirect Cost	60,720	66,405.9	100,812.94	129,566.76	107,952.45	53,350.64	518,808.7
	612,720	670,095.9	1,017,294.24	1,307,446.44	1,089,338.37	538,356.48	5,235,251.43

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	TOTAL
CIAT	612,720	670,095.9	1,017,294.24	1,307,446.44	1,089,338.37	538,356.48	5,235,251.43
	612,720	670,095.9	1,017,294.24	1,307,446.44	1,089,338.37	538,356.48	5,235,251.43

Explanations of these costs in relation to the planned 2022 outcomes

Module 3 establishes big data to hit the ground, and actively solve development problems head on. The total budget over 6 years is US\$5.24m. This budget progressively grows from Year 1 to a maximum in Year 4 as per the Theory of Change whereby the Inspire projects play a more significant role in the Platform in later years. In year 1, a total of five 100k Pilot Inspire grants are budgeted. In year 2, an additional three Pilot Inspire projects are budgeted, and one Scaled Inspire project of US\$250k which will be selected from the best of the crop from Year 1. In Year 3, two additional Pilot Inspire projects are budgeted, plus the second year of the Scaled Inspire project from Year 2, plus one additional new Scaled Inspire project. In year 4, two additional Pilot Inspire projects are budgeted, plus two new Scaled Inspire projects and second year funding from the Year 3 Scaled Inspire project. Year 5 consists of funding for the 5th and final Scaled Inspire project, and second year funding for the two Scaled Inspire projects of year 4. The final year of the Platform consists of 250k funding for the final year of the final Scaled Inspire project. Impact assessment is budgeted at US\$158k per year from year 3 onwards. US\$75k/yr of this is to cover economist staff, and US\$83k/yr for operations (travel, survey instruments, field costs). 10% of Andy Jarvis time is budgeted in this Module to co-lead with the Platform Coordinator. Supplies and services related to managing the Inspire project calls, contracting and reporting amount to US\$25k per year, with an additional US\$5k budgeted for travel to document Inspire projects in action.

50% of this budget is assumed to go to non-CGIAR partners, and 50% to CGIAR partners. The true distribution of this will only become apparent once initial projects are selected and contracted, but criteria will require non-CGIAR partners to have an active, funded role in the projects. In the meantime, this budget is assigned to CIAT who will manage the sub-grants.

Budget under-execution risks will depend on the speed by which CIAT can issue contracts to Inspire projects. A standard, simple template will be developed and issued to all grantees in the hope that this speeds up administrative processes.

3.3.2.3 Additional explanations for certain accounting categories

Benefits:

National Research Staff (NRS): Fringe benefits for national staff (costs for all benefits are added to the base salary to provide the total cost of the position) are comprised of legal benefits (local mandatory) and extralegal benefits (CIAT mandatory) and the provisions to cover local legal requirements such as: Pension - social security, training and development, occupational health, transportation costs and subsidies, work clothes and personnel protection requirement, and food subsidy.

International Research Staff (IRS): Fringe benefits for international staff (costs for all benefits are added to the base salary to provide the total cost of the position) are comprised of housing allowance, education allowance, car allowance, Cost Of Living Allowance (COLA), hardship, home leave tickets, insurance, retirement contribution, occupational health, training and development, repatriation and relocation provisions.

Other supplies and services: Discussed above

3.3.2.4 Other Sources of Funding for this Project

Opportunities exist for co-funding the Inspire calls with other actors in the big data community and other donors interested in developing big data innovations. The Platform Secretariat will aggressively seek such funding in order to issue Inspire project calls jointly. CIAT has had significant success in fundraising for big data related activities, and has close partnerships with WorldBank, Government of Korea, Government of Colombia amongst others who will be approached to provide bilateral support. CRPs themselves will also be approached to provide some co-funding if resources permit. This will potentially reduce the W1+2 burden of this module in Years 2-5. In year 1, seed W1+2 funding will be critical to generate a successful first portfolio of innovations.

3.3.2.5 Budgeted Costs for certain Key Activities

	Estimate annual average cost (USD)	Please describe main key activities for the applicable categories below, as described in the guidance for full proposal
Gender	300,000	Some Inspire project topics will be targeted specifically to address gender gaps, and innovate around big data and ICT means of reaching marginalised women farmers and female headed households. Additionally, Inspire projects on other topics will examine gender-related impacts of the innovation, and impact assessment studies will explicitly examine the gender-related impacts that big data innovations have.
Youth	200,000	Similarly to gender, some Inspire project topics may specifically address youth related impacts. Given the technological nature of big data innovations, it is to be expected that youth are more likely to engage in projects given their additional capacity to manage ICTs. Likewise, impact assessment efforts will look at youth related impacts of different innovations.
Capacity development	500,000	This module is almost entirely focused on building innovation capacity in CGIAR and with partners, hence a heavy load of budget can be considered aligned to capacity building.
Impact assessment	158,000	Impact assessments of a range of big data related innovations for agricultural development are explicitly budgeted in years 3-6. Survey instruments and other techniques will be used in a range of countries and projects to systematically assess the impacts of scaled Inspire projects.
Intellectual asset management	50,000	Intellectual assets resulting from Inspire projects will be managed as per IA guidelines by project leaders, with support from the Platform Secretariat.
Open access and data management	300,000	A significant part of projects in the Inspire module are expected to generate data and manage it in accordance with the OAODM policy of CGIAR.
Communication	100,000	Significant communication activities will be made around Inspire projects from the Secretariat, but also by project leaders themselves, in order to disseminate good practice in big data related to agricultural development.

3.3.2.6 Other

Further budget information available in Annex 5.

3.3.3 Module Uplift Budget

2022 ADDITIONAL OUTPUTS/SERVICE DESCRIPTION	AMOUNT NEEDED (\$)	W1 + W2 (%)	W3 (%)	BILATERAL (%)	OTHER (%)
11. SIGNIFICANTLY EXPANDED INSPIRE PROJECT FUNDING AND IMPACT ASSESSMENT Under this uplift scenario, the budget of the Inspire model is doubled, permitting: • Inspire projects to be funded to the tune of US\$200k instead of US\$100k • Scaled Inspire projects to be funded to US\$500k instead of US\$250k • Vastly increased impact assessment and further reaching synthesis products The entirety of this uplift budget is not needed to expand the Inspire efforts in Module 3, hence any additional support received will gradually increase the projects budgets that can be supported. Bilateral and W3 support received for this may also offset the additional budget request.	5,235,000	100	0	0	0

4. ANNEXES

1. **Business case: Expertise and track record –CV's..**
2. **Business case: Interactions between the Platform and users**
3. **Open Access (OA) and Open Data (OD) Management**
4. **Intellectual Asset Management (IA Management**
5. **Other Annexes - Budget narratives**
6. **Other Annexes – References**
7. **Open Access (OA) and Open Data (OD) Management**
8. **Other Annexes - Consultations**

4.1 ANNEX 1 EXPERTISE AND TRACK RECORD - CVS

Elizabeth ARNAUD

Current Position and Affiliation

- Scientist for Agricultural Ontologies, Bioversity International, France

Profile

- Data management and communication, research interest in agrobiodiversity data & information; Open Data standards for agriculture, semantics, fieldbooks

Employment

- 2008- to date Scientist IRS, Project leader – Bioversity International, Italy and France
- 1998 – 2008 Bioversity International, Coordinator of Musa Germplasm Information System
- 1990- 1998 Genetic Resources Information specialist, International Network for the Improvement of Banana and Plantains (INIBAP)
- 1987-1989 Consultant, Communication & publications Dept, International Rice Research Institute (IRRI), Philippines

Education

- M.Sc. In biology , University of Paris-Orsay, France
- M.Sc. In Scientific data management, Ecole National des Arts et Metiers, Paris, France

Recent and Relevant Peer-Reviewed Publications

- Cooper, L., et al. (2013). The Plant Ontology As A Tool For Comparative Plant Anatomy And Genomic Analyses. *Plant and Cell Physiology* 54(2): e1(1–23) doi:10.1093/pcp/pcs163
- Thorman, I., et al.(2012). Digitization and online availability of original collecting mission data to improve data quality and enhance the conservation and use of plant genetic resources. *Gen. Res. Crop Evol.* 59(5):634-644. DOI: 10.1007/s10722-012-9804-z
- Arnaud, E., et al. (2012). Towards a Reference Plant Trait Ontology for Modeling Knowledge of Plant Traits and Phenotypes. In: *Proc. International Conference on Knowledge Engineering and Ontology Development*. SciTePress. DOI: 10.5220/0004138302200225. Pp.220-225.
- Shrestha, R., et al. (2012). Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. *Frontiers in Plant Physiology* 3 Article 326: doi: 10.3389/fphys.2012.00326 , ISSN: 1664-042X

Other Evidence of Leadership

- Coordinator of the *Musa* Germplasm Information System (MGIS) for 22 collections
- Leader of the [Crop Ontology](#) project of the Generation Challenge ProgrammeChair of the GBIF/Bioversity Task Group on GBIF data fitness for use for Agrobiodiversity

Role in Big Data Platform

Development of the semantics (Module 1), Agrobiodiversity data standards

Carlo AZZARRI

Current Position and Affiliation

- Research Fellow, Environment and Production Technology Division (EPTD), International Food Policy Research Institute (IFPRI), Washington, DC, USA

Profile

- Quantitative micro and macroeconomic level analysis of the relationship between poverty, nutrition, food security, agriculture, and migration.
- Extensive experience in household survey microdata analysis, implementation and supervision, having conducted fieldwork activities in Africa south of the Sahara, Latin America, and Eastern Europe.
- Managing large multi-topic and specialized household survey databases

Employment

- 2011-Present, Research Fellow, International Food Policy Research Institute, USA
- 2006-2011, Short- and Extended-term Consultant at the Development Economics Research Group, World Bank, USA
- 2003-2006, Consultant, Economics and Social Department, FAO, Italy

Education

- 2005, Ph.D. in Economics, University of Rome “Tor Vergata”, Italy
- 2000, M.Phil. in Economics, University of Rome “Tor Vergata”, Italy

Recent and Relevant Peer-Reviewed Publications

- Azzarri, C., et al. (2016). Subnational socio-economic dataset availability. *Nature Climate Change* 6: 115-116. DOI: 10.1038/nclimate2842.
- Azzarri, C. and Cross, E. (2016). Improved spatially-disaggregated livestock measures for Uganda. *The Review of Regional Studies* 46 (1).
- Azzarri, C., et al. (2015). Does Livestock Ownership Affect Animal Source Foods Consumption and Child Nutritional Status? Evidence from Rural Uganda. *Journal of Development Studies* 51(8).
- Davis, B., et al. (2010). A Cross Country Comparison of Rural Income Generating Activities. *World Development* 38(1).

Other Evidence of Leadership

- Team leader of IFPRI M&E Africa Research In Sustainable Intensification for the Next Generation (Africa RISING)
- Responsible for economic analysis in the Spatial Data and Analytics Theme (EPTD)

Role in Big Data Platform

- Socio-economic data harmonization and analysis
- Impact evaluation

Melanie BACOU

Current Position and Affiliation

- Senior Program Manager, Environment Production Technology, International Food Policy Research Institute (IFPRI), Washington, DC, USA

Profile

- Agricultural economist with strong quantitative, analytical, and writing / presentation skills
- Over 15 years of experience deploying and evaluating data-centric and technology-oriented solutions to support economic development and aid programs; Quantitative and spatial analysis, cost/benefit analysis, household surveys, experimental designs, impact evaluation
- R, STATA, GNU Octave, QGIS, GEOS/GDAL, GRASS, MS SQL Server, MySQL, PostGIS
- Experienced developing on many CMS and LMS platforms; Server administration

Employment

- 2014 – present Snr. Program Manager, International Food Policy Research Institute (IFPRI)
- 2007 – 2014 Independent Consultant (Bill and Melinda Gates Foundation, African Development Bank, Development Gateway Foundation, Purdue University)
- 2005 – 2007 Senior Analyst, Abt Associates Inc., Bethesda MD
- 2000 – 2005 ICT/Data Specialist, Center for Global Trade Analysis (GTAP)

Education

- M.Sc. International Trade and Development Economics, 2001
- M.A. International Management (French Diplôme de Grande École, German Diplom Kaufmann), ESCP Europe, 1997

Recent and Relevant Peer-Reviewed Publications

- Azzarri, C., M. Bacou, C.M. Cox, Z. Guo, and J. Koo. 2016. Subnational socio-economic dataset availability. *Nature Climate Change* 6(2): 115–116.
- HarvestChoice, International Food Policy Research Institute (IFPRI); University of Minnesota, 2016, "CELL5M: A Multidisciplinary Geospatial Database for Africa South of the Sahara", <http://dx.doi.org/10.7910/DVN/G4TBLF>, Harvard Dataverse, V2

Other Evidence of Leadership

- CGIAR representative in the Agricultural Investment Data Community of Practice and IATI Agriculture Working Group (developing data standards and tools for the collection of investment-related and scientific data on agriculture)

Role in Big Data Platform

- Agricultural investment mapping and data management
- Big data analytics using R framework

Matthew REYNOLDS

Current Position and Affiliation

- Head of Wheat Physiology, International Maize and Wheat Improvement Center (CIMMYT), Mexico.

Profile

- Genetic and physiological bases of crop adaptation
- Germplasm development and breeding methodology
- Strategy development, external fundraising and project management

Employment

- 1989-Present, Wheat Program, International Maize and Wheat Improvement Center (CIMMYT), Mexico.
- 1984-1989, Department of Fruit & Vegetable Sciences, Cornell University, USA

Education

- 1989, Ph.D. in Horticulture Science, Cornell University, USA.
- 1984, M.Sc. in Crop Physiology, Reading University, UK.
- 1983, B.A. in Botany, Oxford University, UK.

Recent and Relevant Peer-Reviewed Publications

- Reynolds, M.P. and Langridge, P. (2016). Physiological Breeding. *Current Opinions in Plant Biology* 31: 162–171 doi:10.1016/j.pbi.2016.04.005
- Reynolds M.P, et al. (2016). An integrated approach to maintaining cereal productivity under climate change. *Global Food Security* 8: 9-18
- Reynolds, M.P., Tattaris, M., Cossani, C.M., Ellis, M., Yamaguchi-Shinozaki, K., Saint Pierre, C. (2015). Exploring genetic resources to increase adaptation of wheat to climate change. In *Advances in Wheat Genetics: From Genome to Field*. In: Ogihara, Y., Takumi, S and Handa, H. Springer, Japan.
- Valluru, R.; Reynolds, M.P. and Salse, S. (2014). Genetic and molecular bases of yield-associated traits: a translational biology approach between rice and wheat. *Theoretical and Applied Genetics* 127(7):1463-1489.

Other Evidence of Leadership

- Head of wheat physiology, 1996-Present. International Maize and Wheat Improvement Center (CIMMYT), Mexico.
- Leading the foundation of a *Heat and Drought Wheat Improvement Consortium* (HeDWIC), an international partnership involving hundreds of plant abiotic stress experts; target budget of \$50m+.

Role in Big Data Platform

- Leader of CoP on Crop Modelling

Chandrashekhar BIRADAR

Current Position and Affiliation

- Principal Scientist and Head of the Geoinformatics Unit, International Center for Agricultural Research in Dry Areas (ICARDA), Amman, Jordan

Profile

- Time-series analysis for agro-ecosystems research and outreach in dry areas
- Geospatial bigdata analytics for delivering better interventions and technology scaling

Employment

- 2013- Head Geoinformatics Unit and Principal Scientist, ICARDA, Jordan
- 2008-2013 Research Professor/Research Scientist University of Oklahoma, Norman, USA
- 2007-2008 Research Scientist, University of New Hampshire, Durham, NH, USA
- 2004-2007 Remote Sensing Expert and Postdoc, IWMI, Colombo, Sri Lanka Colombia

Education

- 2005, Ph.D., RS and Envi. Sciences, IIRS, ISRO, and Uni. of Pune, India
- 2000, M.Sc., Genetic Engineering, Uni. of H&F, HP, India

Recent and Relevant Peer-Reviewed Publications

- Dong, et al. (2016). Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote Sensing of Environment. [doi:10.1016/j.rse.2016.02.016](https://doi.org/10.1016/j.rse.2016.02.016)
- Ghanem, et al. (2015), Production potential of Lentil (*Lens culinaris* Medik.) in East Africa, Agricultural Systems 137 (2015) 24–38
- Dong, et al. (2015). Comparison of four EVI-based models for estimating gross primary production of maize and soybean croplands and tallgrass prairie under severe drought. *Remote Sensing of Environment*. 162 (2015), 154–168
- Biradar, et al. (2009). Quantifying the area and spatial distribution of double- and triple-cropping croplands in India with multi-temporal MODIS imagery in 2005. *International Journal of Remote Sensing*. 32(2), 367-386
- Biradar, et al. (2009). A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 11 (2) 114–129.

Other Evidence of Leadership

- Thematic leader and focal point for the Geoinformatics Science, Technology and Application (GeSTA)
- PI/Col/Activity leads on a number of spatial data analytics-relevant projects at ICARDA

Role in Big Data Platform

- Spatial data analytics and integration (Module 1)

Medha DEVARE

Current Position and Affiliation

- Data and Knowledge Manager, CGIAR Consortium, Montpellier, France

Profile

- Cropping systems agronomy with experience in developing countries of S. Asia, Africa
- Molecular ecology and bioinformatics with expertise in a variety of molecular and in silico techniques and analyses
- Data and knowledge management, with experience in Semantic Web and interoperability approaches (e.g. ontologies, vocabularies)
- Project management, resource mobilization

Employment 2014-Present, Data and Knowledge Manager, CGIAR Consortium, France

- 2014-Present, Data and Knowledge Manager, CGIAR Consortium, France
- 2010-2014, Cropping Systems Agronomist, International Maize and Wheat Improvement Center, Mexico (based in Nepal)
- 2004-2010, Bioinformatics Specialist, Cornell University, USA

Education

- 2000, Ph.D., Crop and Soil Sciences, Cornell University, USA
- 1993, M.S., Environmental Toxicology, Cornell University, USA

Selected Recent and Relevant Peer-reviewed Publications

- Xue K., R.C. Serohijos, M. Devare, J. Duxbury, J. Lauren, and J.E. Thies. 2012. Short-term allocation and root lignin of Cry3Bb Bt and nonBt corn in the presence of corn rootworm. *Appl Soil Ecology* 57: 16-22.
- Xue, K., R.C. Serohijos, M. Devare, and J.E. Thies. 2011. Decomposition rates and residue-colonizing microbial communities of *Bacillus thuringiensis* insecticidal protein Cry3Bb-expressing Cry3Bb (Bt) and Non-Bt corn hybrids in the field. *Applied and Environmental Microbiology* 77(3): 839-846.
- Krafft, D.B., N.A. Cappadona, B. Caruso, J. Corson-Rikert, M. Devare, B.J. Lowe. 2010. VIVO: Enabling National Networking of Scientists. *Web Science Conf.* April 26-27, 2010, Raleigh, NC, USA.
- Lu, H., et al. (2010). Decomposition of Bt transgenic rice residues and response of soil microbial community in rapeseed–rice cropping system. *Plant Soil* 336: 279-290.

Other Evidence of Leadership

- PI on CGIAR Open Access, Open Data initiative; co-PI on Virtual Information Platform
- PI on CSISA-Nepal project at CIMMYT; co-PI on VIVO semantic web project

Role in Big Data Platform

- Lead, Open Access and Open Data; Module 1

Zhe GUO

Current Position and Affiliation

- Senior Geographic Information System Coordinator, International Food Policy Research Institute (IFPRI), Washington, DC, USA

Profile

- Spatiotemporal data analysis with data mining and machine learning technique
- Land classification and monitoring, image interpretation
- Computer programming with Python, Matlab, and C

Employment

- 2014-Present, Senior GIS coordinator, International Food Policy Research Institute, USA
- 2010-2014, GIS coordinator, International Food Policy Research Institute, USA
- 2007-2010, Research Analyst, International Food Policy Research Institute, USA

Education

- 2007, M.A. in Geography, University of Maryland, College Park, USA
- 2002, M.S. in Ecology, Chinese Academy of Science, China

Recent and Relevant Peer-Reviewed Publications

- Azzarri, C., et al. (2016). Subnational socio-economic dataset availability. *Nat. Clim. Chang.* 6(2): 115–116.
- Guo, Z. (2014). Map teff in Ethiopia: An approach to integrate time series remotely sensed data and household data at large scale. *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International* 1:2138-2141.
- You, Liangzhi, et al. (2014). Irrigation potential and investment return in Kenya. *Food Policy* 47 (August 2014): 34-45.
- Guo, Z. and Hou, Y. (2002). Establishment and Application of Expert System of Ecological Balanced Expert Consultation System. *Ecological model and parameter system. Chinese Journal of Soil Science* 24:762-774.

Other Evidence of Leadership

- IFPRI's Representative of Consortium for Spatial Information in CGIAR Centers

Role in Big Data Platform

- Geospatial data harmonization and analytics
- Community of Practice on the geospatial data and analysis

Glenn HYMAN

Current Position and Affiliation

- Senior Staff in Land Use and Geographic Information Science, International Center for Tropical Agriculture (CIAT), Cali, Colombia

Profile

- Temporal and spatial land use change dynamics
- Targeting technology, geographic dimensions of tropical agriculture

Employment

- 1998-Present, Research Fellow, International Center for Tropical Agriculture, Colombia
- 1996-1998, Postdoctoral Fellow, International Center for Tropical Agriculture, Colombia

Education

- 1997, Ph.D., Geography, University of Tennessee, USA
- 1990, M.A., Geography, Appalachian State University, USA
- 1985, B.A., International Studies, University of North Carolina, USA

Recent and Relevant Peer-Reviewed Publications

- Imbach, P., M. Manrow, E. Barona, A. Barretto, G. Hyman, and P. Ciais. 2015. Spatial and temporal contrasts in the distribution of crops and pastures across Amazonia: A new agricultural land use data set from census data since 1950, *Global Biogeochem. Cycles*.
- Börner, J., S. Wunder, S. Wertz-Kanounnikoff, G. Hyman & N. Nascimento. 2014. Forest law enforcement in the Brazilian Amazon: Costs and income effects. *Global Environmental Change*. 29:294-305.
- Hyman, G., D. Hodson and P. Jones. 2013. Spatial analysis to support geographic targeting of genotypes to environments. *Frontiers in Physiology*. 4:40.
- Hyman, G., S. Fujisaka, P. Jones, S. Wood, C. de Vicente and J. Dixon. 2008. Strategic approaches to targeting technology generation: Assessing the coincidence of poverty and drought-prone crop production. *Agricultural Systems*. 98:50-61.
- Hyman, G., C. Larrea and A. Farrow. 2005. Methods, results and policy implications of poverty and food security mapping assessments. *Food Policy*. 30(5-6):453-460.

Other Evidence of Leadership

Founding coordinator of CGIAR Consortium for Spatial Information (CGIAR-CSI).

- PI on agricultural evaluation repository (AgTrials), CGIAR poverty mapping initiative, research and development priority setting initiatives.

Role in Big Data Platform

- Co-leader WG on Data Products

Andy JARVIS

Current Position and Affiliation

- FP2 Leader, CCAFS and Research Area Director, Decision and Policy Analysis, International Centre for Tropical Agriculture (CIAT), Cali, Colombia

Profile

- Research to support the goals of alleviating poverty and protecting essential ecosystem services

Employment

- 2011–present, Flagship 2 Leader, CGIAR Research Program on Climate Change, Agriculture and Food Security, Cali, Colombia
- 2012–present, Research Area Director, Decision and Policy Analysis, International Centre for Tropical Agriculture (CIAT), Cali, Colombia
- 2009–2010, Theme Leader, CGIAR Challenge Program on Climate Change, Agriculture and Food Security
- 2008–2011, Program Leader, Decision and Policy Analysis, International Centre for Tropical Agriculture (CIAT), Cali, Colombia

Education

- Ph.D., Department of Geography, King's College, London, UK (2005)
- M.S., Department of Geography, King's College, London, UK (2001)

Selected Recent and Relevant Peer-reviewed Publications

- Khoury, C.K., et al. (2014). Increasing homogeneity in global food supplies and the implications for food security. PNAS, 2014 DOI: 10.1073/pnas.1313490111
- Lipper, L., et al. (2014). Climate-smart agriculture for food security. Nature Climate Change 4: 1068–1072. doi:10.1038/nclimate2437.
- Vermeulen, S.J., et al. (2013). Addressing uncertainty in adaptation planning for agriculture. Proceedings of the national academy of sciences of the United States of America 110 (21): 8357-8362.
- Jarvis, A., et al. (2012). Is Cassava the Answer to African Climate Change Adaptation? Tropical Plant Biology 5(1): 9-29, doi:10.1007/s12042-012-9096-7.

Other Evidence of Leadership

- PI on a US\$15m agreement with the Ministry of Agriculture and Rural Development of Colombia, a NERC-ESPA funded project on ecosystem services for food and nutritional security in the Amazon, and has successfully led for CCAFS a Theme/Flagship for 5 years.
- CIAT and CCAFS, managed resources >US\$20m per year, delivering development outcomes in big data in extension, inclusive business models, and influencing CSA investment.

Role in Big Data Platform

- Co-leader

Daniel JIMÉNEZ R.

Current Position and Affiliation

- Scientist- leader of the expert group on Big Data, International Center for Tropical Agriculture (CIAT), Cali, Colombia

Profile

- Agronomist and data scientist
- Data mining, empirical modelling approaches to agronomy, closing yield gaps, using resources more efficiently , gaining competitive strength, and coping with climate change

Employment

- 2013-Present, Senior Scientist, International Center for Tropical Agriculture, Cali, Colombia
- 2010-2013, Postdoctoral Fellow, International Center for Tropical Agriculture, Cali, Colombia.

Education

- 2013, Ph.D. in Applied Biological Sciences: Agricultural Science. Ghent University. Belgium
- 2002, Ing. Agr., Universidad de Caldas, Manizales, Colombia.

Recent and Relevant Peer-Reviewed Publications

- Jiménez, D., et al. (2016). from Observation to Information: Data-Driven Understanding of on Farm Yield Variation. PloS One 11, no. 3 : e0150015.
- Jiménez, D., et al. (2011). Interpretation of commercial production information: a case study of lulo (*Solanum quitoense*), an under-researched Andean fruit. Agricultural Systems 104(3):258–270.
- Jiménez, D., et al. (2009). Analysis of Andean blackberry (*Rubus glaucus*) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data. Computers and Electronics in Agriculture 69(2):198–208.
- Satizábal, et al. (2012). Enhancing decision-making processes of small farmers in tropical crops by means of machine learning models. In: Bolay J-C; Schmid M; Tejada G; Hazboun E. (eds.). Technologies and innovations for development – Scientific cooperation for a sustainable future. Paris: Springer. Pp. 265–277.

Other Evidence of Leadership

- PI on several projects in Colombia, Nicaragua, Peru, Argentina, Uruguay on Big Data and Site-specific agriculture in a wide range of crops (> 3M USD)
- Development of and support to the implementation of several projects on adaptation to climate change.

Role in Big Data Platform

- CoP leader on data driven agronomy

Jawoo KOO

Current Position and Affiliation

- Research Fellow, Spatial Data and Analytics Theme Leader, International Food Policy Research Institute (IFPRI), Washington, DC, USA

Profile

- Dynamic simulation of food production systems on geospatial, grid-based modeling platform at regional and global scales – primarily using the DSSAT Cropping System Model
- Managing large agricultural, geospatial research datasets
- Computer programming with Java, Fortran, JavaScript, and PHP

Employment

- 2008-Present, Research Fellow, International Food Policy Research Institute, USA
- 2007-2008, Postdoctoral Fellow, International Food Policy Research Institute, USA

Education

- 2007, Ph.D., Agricultural & Biological Engineering, University of Florida, USA
- 2002, M.S., Agricultural & Biological Engineering, University of Florida, USA

Selected Recent and Relevant Peer-reviewed Publications

- Azzarri, C., et al. (2016). Subnational socio-economic dataset availability. *Nat. Clim. Chang.* 6(2): 115–116.
- Nkonya, E., et al. (2016). Global Cost of Land Degradation. *In: Nkonya, E., Mirzabaev, A., von Braun, J. (eds.), Economics of Land Degradation and Improvement – A Global Assessment for Sustainable Development.* Cham, Switzerland: Springer International Publishing. Pp. 117–165.
- Rosegrant, M.W. Et al. (2014). Food security in a world of natural resource scarcity: Role of agricultural technologies. Washington, D.C.: IFPRI.
- White, J.W., et al. (2013). Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Comput. Electron. Agric.* 96:1–12.

Other Evidence of Leadership

- Serving as the Spatial Data and Analytics Theme Leader at IFPRI and the Global Coordinator of the CGIAR Consortium for Spatial Information (CGIAR-CSI)
- PI on a number of data analytics-relevant projects at IFPRI: HarvestChoice, Virtual Information Platform, and Technology Platform

Role in Big Data Platform

- Co-leader

Gideon KRUSEMAN

Current Position and Affiliation

- Ex-ante & foresight specialist and Big Data focal point
- International Maize and Wheat Improvement Center (CIMMYT), El Batán, Mexico

Profile

- Expert in quantitative economic and bio-economic modeling of complex systems
- Expert in quantitative and qualitative ex-ante and ex-post policy analysis, impact assessment, monitoring and evaluation
- Expert on high variety and high complexity Big Data and data integration

Employment

- 08/2015 to date: Ex-ante and foresight specialist at CIMMYT
- 03/2015 to 08/2016 Consultant Big Data
- 03/2006 to 03/2015: Lead on environmental economic modelling, at LEI Wageningen, the Netherlands
- 03/2004 to 03/2006: Senior Research fellow at the Institute of Environmental studies in Amsterdam, the Netherlands

Education

- 2000, Ph.D., Development economics, Wageningen University
- 1988, M.Sc, Development economics, Wageningen University

Selected Recent and Relevant Peer-reviewed Publications

- Gebrezgabher, S., G. Kruseman; M.P Meuwissen, D. Lakner, A.G Oude Lansink (2015) Factors influencing adoption of manure separation technology in the Netherlands. Journal of Environmental Management 150: 1-8.
- Hagos, F.; Yawez, E.; Yohannes, M.; Mulugeta, A.; Abraha, G.G.; Abreha, Z.; Kruseman, G.; Linderhof, V.G.M. (2013) Small scale water harvesting and household poverty in Northern Ethiopia. in: Nature's Wealth: The Economics of Ecosystem Services and Poverty, Papyrakis, E., Bouma, J., Brouwer, R., Beukering, P.J.H. van. (editors) - Cambridge : Cambridge University Press, pp. 265 - 282
- Kabubo-Mariara, J.; Linderhof, V.G.M.; Kruseman, G.; Atieno, R.; Mwabu, G. (2010) Poverty-environmental links: The impact of soil and water conservation and tenure security on household welfare in Kenya. Journal of Development and Agricultural Economics 2 (1). - p. 041 - 053.

Other Evidence of Leadership

- Research leader on ex-ante and foresight studies at CIMMYT
- PI and/or project leader/director on numerous data analytical projects at various research organizations

Role in Big Data Platform

- Lead CoP on socioeconomic data

Ricardo LABARTA

Current Position and Affiliation

- Senior Scientist, Leader of Impact Assessment Research Group, International Center for Tropical Agriculture (CIAT), Cali, Colombia

Profile

- Impact assessment and outcome evaluations
- Socio-economic research
-

Employment

- 2013-present Senior Scientist, International Center for Tropical Agriculture (CIAT), Cali, Colombia
- 2006-2013, Regional Economist for Sub-Saharan Africa. International Potato Center (CIP)
- 2001-2005, Graduate Research Assistant, Department of Agricultural, Food & Resource Economics, Michigan State University
- 1998-2001, Research Associate, International Center for Tropical Agriculture (CIAT), Pucallpa, Peru

Education

- Ph.D. in Environmental & Natural Resource Economics, Michigan State University, 2005.
- M.S. in Agricultural Economics, Michigan State University, 2004

Selected Recent and Relevant Peer-reviewed Publications

- Yamano, T., A. Aminou, R. Labarta, Z.M. Huelgas and S. Mohanty (2016). Adoption and impacts of international rice research. *Global Food Security* 8:1-8.
- Larochelle, C., Alwang, J., Norton, G. W., Katungi, E. and Labarta, R.A. (2015). Impacts of improved bean varieties on poverty and food security in Uganda and Rwanda, In T.S. walker and J. Alwang (eds.) *Crop improvement, adoption and impact of improved varieties in food crops in sub-Saharan Africa*. Wallingford, UK: CABI Publisher. Pp 314-337.
- Scott, G., Labarta, R. and Suarez, V. (2013). Benchmarking food crop diversity in southern Africa: The case of potatoes and potatoes products 1961-2010. *American Journal of Potato Research* 90(6):497-515.
- Scott, G., Labarta, R. and Suarez, V. (2013). Booms, busts, and emerging markets for potatoes in East and Central Africa 1961–2010. *Potato Research* 56:205-236.

Other Evidence of Leadership.

- Managed research teams in CIAT (17 staff members) and CIP (10 staff members) and supervised graduate students (6 PhD and 11 MsC)
- Led research partnerships each consisting on 3-6 partners in Colombia, 2015 to date, Rwanda, 2014 to date, Honduras 2013 to date, Uganda 2010-2013, Ethiopia 2010-2013, Mozambique 2006-2010
- Led fundraising to investigate impacts of water harvesting technologies (US\$987,000), impacts of cassava research (US\$ 553,000) impacts of sustainable agriculture (US\$ 458,000) and impacts of genetic crop improvement (US\$ 700,000)

Role in Big Data Platform

- Module 3, Impact assessment evaluation

Charles MCKEOWN

Current Position and Affiliation

- Research Faculty, Department of Agricultural Food and Resource Economics, East Lansing Michigan, United States

Profile

- Decision support, geospatial, and outreach specialist.
- Efficiency and renewable energy analysis, resource assessment, cluster analysis, ecological assessments.
- Network Collaboration building and enhancing Communities of Practice.
- International experience in Tanzania, Nigeria, Rwanda, Djibouti, Myanmar, Colombia

Employment

- 2011–Current, Research and outreach faculty member, Department of Agriculture Food and Resource Economics, Michigan State University.
- 2012–Current, McKeown Technical Services. I provide consulting services focused on strategic planning, asset based economic development strategies and renewable energy development.
- 2012–2015, Assistant Director, Global Center for Food Systems Innovation, Michigan State University, East Lansing Michigan,
- 2005–2011, Associate Director for Decision Support, Land Policy Institute, Michigan State University, East Lansing Michigan.

Education

- M.S. (Hons) in Ecology, Evolutionary Biology and Behavior, Michigan State University, 2003.
- B.S. (Hons) in Entomology, Michigan State University, USA, 2001.

Recent and Relevant Peer-Reviewed Publications

- McKeown, C., et al. (2009). Final Report of the Michigan Wind Resources Zone Board. Michigan Public Service Commission Report, Michigan State University, East Lansing, Michigan.
- McKeown, C., et al. (2011). On development of a Wind Prospecting Tool for Industry and Policy Decision Support. Energy Policy, December 2010.
- Adelaja, A, McKeown, C. (2012). Assessing Offshore Wind Potential. Energy Policy, January 2012.
- McKeown, C. (2014). Asset Based Development and Decision Support System for Nigeria. Report presented to the Nigerian President's Economic Management Team Conference, Abuja, 2015

Other Evidence of Leadership

- Member of MSU team that Secured, Lunched and Managed GCFSI, \$32 Million USD USAID HESN Lab.
- Founding Director of the MSU Land Policy Institute, formed with \$14.2 million USD grant from the W.K. Kellogg Foundation

Role in Big Data Platform

- Expert in big data policy research, and support in Module 2 Convene

Steven PRAGER

Current Position and Affiliation

- Senior Scientist for Integrated Modeling, International Center for Tropical Agriculture (CIAT), Cali, Colombia

Profile

- Ex-ante modeling of economic impacts of climate change
- Dynamic spatial modeling using evolutionary algorithms and complex network theory
- Python programming, large database management, quantitative methods

Employment

- 2014-Present, Senior Scientist, International Center for Tropical Agriculture, Colombia
- 2004-2014, Assistant and Associate Professor, Department of Geography, University of Wyoming, USA

Education

- 2002, Ph.D., Geography, Simon Fraser University, Canada
- 1995, M.A., Geography, University of North Carolina-Charlotte, USA

Recent and Relevant Peer-Reviewed Publications

- Jiménez, D., et al. (2016). From Observation to Information: Data-Driven Understanding of on Farm Yield Variation. PloS One 11, no. 3 (2016): e0150015.
- Prager, S.D. and Pfeiffer, C. (2015). Social-ecological networks for understanding water management and spatial resilience. Ecology & Society 20(4):13.
- Mbügwa, G.W. and Prager, S.D. (2015). Utilization of spatial decision support systems decision-making in dryland agriculture: A Tifton burr-clover case study. Computers and Electronics in Agriculture 118:215-224.
- Prager, S.D. and Wiegand, R.P. (2014). Modeling use of space from social media data using a biased random walker. Transactions in GIS 18(6):817-833.

Other Evidence of Leadership

- Served as President and member of the Board of Directors for the University Consortium for Geographic Information Science
- PI on large analytics projects in both academic and CGIAR settings.

Role in Big Data Platform

- Co-leader WG on Data Products

Lisa-Maria REBELO

Current Position and Affiliation

- Core Theme Leader, Decision Analysis and Information Systems, Water Land and Ecosystems Program; Senior Researcher, remote sensing and GIS, International Water Management Institute, Lao P.D.R.

Profile

- Development of remote sensing tools for water management practices on wetlands
- Development of remote sensing methodologies for basin scale hydrological and ecological functioning and water availability and allocation

Employment

- 2014 – present, Core Theme Leader, Water Land and Ecosystems Program, Lao P.D.R.
- 2013 – present, Senior Researcher (Remote Sensing and GIS), International Water Management Institute, Lao P.D.R.
- 2009 – 2013, Researcher (Remote Sensing and GIS), International Water Management Institute, Ethiopia
- 2006 – 2009, Post Doctoral Fellow, International Water Management Institute, Colombo, Sri Lanka

Education

- Ph.D. in Department of Geography, University College London, UK (2006)
- M.Sc. in Department of Geography, University College, London, UK (2001)

Recent and Relevant Peer-Reviewed Publications

- Fluet-Chouinard, E., B. Lehner, L.-M. Rebelo, et al., (2015). Development of a global inundation map at high spatial resolution from topographic downscaling of coarse scale remote sensing data, *Remote Sensing of Environment* 158:348-361
- Lucas, R., Rebelo, L-M., et al, (2014). Contribution of L-band SAR to Systematic Global Mangrove Monitoring, *Marine and Freshwater Research*, 65:589-603
- Rebelo, L-M., Johnston, R., and McCornick, P.G., (2014). Determining the Dynamics of Agricultural Water Use: Cases from Asia and Africa, *Journal of Contemporary Water Research and Education on Water Data*, 153:79-90
- Lloyd, C.R., Rebelo, L-M., Finlayson, C.M., 2013. Providing low-budget estimations of carbon sequestration and greenhouse gas emissions in agricultural wetlands, *Environmental Research Letters* (8) 015010

Other Evidence of Leadership

- Principal Investigator under the Japanese Aerospace and Exploration Agency's (JAXA) Kyoto and Carbon Protocol Initiative, an international science team
- Coordinator of the Wetlands Theme of JAXA's International Science under the Kyoto and Carbon Initiative team

Role in Big Data Platform

- Spatial data analytics and integration, WG on spatial data

Peter WENZL

Current Position and Affiliation

- Incoming Genetic Resources Program Leader, CIAT, Colombia

Profile

- Characterization of genetic resources to identify and mobilize novel variation that accelerates genetic gains in breeding programs
- Information management at the interface between genebanks and breeding programs
- Genotyping service provision; configuration of genotyping assays for a variety of purposes

Employment

- 2015-2016, DivSeek Liaison, Global Crop Diversity Trust, Germany
- 2010-2014, Leader, Seeds of Discovery (SeeD) Project, CIMMYT, Mexico
- 2009-2010, Manager, Crop Informatics Team, CIMMYT, Mexico
- 2002-2009, Principal Scientist, Diversity Arrays Technology (DART) Pty. Ltd., Australia

Education

- Ph.D, Plant Physiology & Genetics, University of Vienna, Austria (2000)
- M.Sc, Biochemistry and Biophysics, University of Vienna, Austria (1993)

Selected Recent and Relevant Peer-reviewed Publications

- Sehgal D, Vikram P, Sansaloni CP, Ortiz C, Pierre CS, Payne T, Ellis M, Amri A, Petroli CD, Wenzl P, Singh S (2015) Exploring and mobilizing the gene bank biodiversity for wheat improvement. PLoS ONE 10: e0132112
- McCouch S, et al. (2013) Feeding the future. Nature 499: 23-24
- Bedo J, Wenzl P, Kowalczyk A, Kilian A (2008) Precision-mapping and statistical validation of quantitative trait loci by machine learning. BMC Genetics 9, 35
- Wenzl P, Raman H, Wang J, Zhou M, Huttner E, Kilian A (2007) A DART platform for quantitative bulked segregant analysis. BMC Genomics 8, 196
- Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinjohs A, Kilian A (2004) Diversity arrays technology (DART) for whole- genome profiling of barley. PNAS 101, 9915–9920

Other Evidence of Leadership

- Led SeeD project team of 15 IRS and ca. 70 staff overall – SeeD systematically identifies and mobilizes useful genetic variation in genebanks into maize and wheat breeding programs
- Substantially contributed to the start-up and success of a biotech company that provides high-throughput genotyping and informatics services to plant breeders and researchers
- Emerging Leaders Training Program, Thunderbird School of Global Management, Arizona State University, Phoenix

Role in Big Data Platform

- Interface to genebanks and 'genomics-of-genebanks' projects

4.2 ANNEX 2 BUSINESS CASE: INTERACTIONS BETWEEN THE PLATFORM AND USERS

User	Types of interactions
CRPs	Close collaboration with CRP data managers to provide support, tools, and training to ensure that CRP data products are appropriately managed, archived and made available openly to users
	Co-development of Inspire project ideas and funding to pilot big data application in agricultural research and development
	Participation in Communities of Practice and Working Groups, when relevant to the CRP
	Access to shared infrastructure, including common infrastructure and analytics platform, cloud storage and backup, high-end analytics capacity and processing infrastructure
	Platform Secretariat can facilitate access to big data partners with specific expertise that may be needed by CRPs
	Capacity building opportunities for mutually identified knowledge gaps
	Participation in the Big Data convention for interested researchers
	Facilitated access to CGIAR data products, including interoperability opportunities with CRP-developed data platforms
Centres	Close collaboration with Center data managers to provide support, tools, and training to ensure that CRP data products are appropriately managed, archived and made available openly to users
	Targeted funding to Center data management teams to meet OAODM policy compliance
	Capacity building opportunities for mutually identified knowledge gaps
	Access to shared infrastructure, including common infrastructure and analytics platform, cloud storage and backup, high-end analytics capacity and processing infrastructure
	Platform Secretariat can facilitate access to big data partners with specific expertise that may be needed by CRPs
	Participation in the Big Data convention for interested researchers
Genetic Gains Platform	Access to processing infrastructure for genetic data and bioinformatic analysis
Genebank Platform	Environmental datasets to add value to germplasm passport information; contribution to concept of Digital Genebank

4.3 ANNEX 3. OPEN ACCESS (OA) AND OPEN DATA (OD) MANAGEMENT

Planning for and Implementing OA/OD in Accordance with the CGIAR OADM Policy and FAIR Principles, including Critical Issues and Anticipated Challenges

The FAIR principles and the Policy broadly require that CGIAR’s information products be discoverable, accessible, interoperable and reusable—all of which are primary goals for this Big Data platform. As indicated by a summary of data reported by 14 Centers through a Phase I OA/OD assessment across CGIAR (Table 1), all reporting Centers and the CRPs led by them have data and publications repositories in place, but these contain highly variable fractions of publications and data that are downloadable without restrictions. A key deliverable of Module 1 (“Organize”) is to ensure that all final publications and associated data from 2014 onward is open by 2018 (as per the terms of the OADM)—as well as high value, low risk, low cost data and publications identified via a data prioritization framework that is already in early draft stage. In brief, the top challenges to compliance with the Policy and FAIR principles relate to this remaining an unfunded mandate, absence of a data-sharing culture, and lack of researcher incentivization and sensitization.

Table 1 Preliminary Center report on publications and data downloadable without restrictions in CGIAR repositories (2015)

Center	Publications			Data		
	In repository	Downloadable, no restrictions	% downloadable	In repository	Downloadable, no restrictions	% downloadable
1	620	0	0	320	0	0
2	293	104	35	11	5	45
3	1599	921	58	16	16	100
4	2863	2315	81	108	55	51
5	2000	33	2	12	12	100
6	400	211	53	523	84	16
7	1859	1859	100	275	187	68
8	8000	4000	50	2539	2539	100
9	4000	1000	25	130	130	100
10	6422	1000	16	0	0	0
11	1500	800	53	32	21	66
12	5525	224	4	1756	415	24
13	7056	759	11	3887	1071	28
14	1195	545	46	65	65	100

These will be addressed across all 3 Modules and through dedicated startup funding to help Centers improve adherence to “open”; the expectation is that OA/OD costs are fully costed in Center programs and projects from 2018. The minimum accomplishment for Module 1 is compliance with CGIAR’s OADM, through funding to Centers, capacity development, sensitization, and support (including guidance on standards, data harmonization and metadata, etc.). Module 2 (“Convene”) enables new partnership networks and models developed with upstream and downstream partners, from public and private sectors will deepen and widen CGIAR’s capacity on big data analytics using CGIAR’s open data—providing a test ground for their validation. Finally, Module 3 will operationalize CGIAR and partner capacity to innovate around big data, with the goal of tackling complex problems with new data-driven approaches, and delivering scalable pilots of effective big data for development solutions. This module will provide further impetus and incentivization for open CGIAR resources.

Note: An indicative data management plan (DMP) is recommended (though not required) by the CRP guidelines; this is not applicable to this platform.

Technical Considerations and Operations

Target 2.3 of Module 1 states that all CGIAR Centers will adopt interoperable data management standards. Activities will focus on syntactic interoperability of repositories (via the use of standards such as XML and SQL, critical to data exchange among machines), as well as semantic interoperability through repositories that support the OAI-PMH protocol at minimum, but ideally are compatible with SPARQL, O-data, and JSON to enable unambiguous data exchange, further allowing machine-computable logic, inference, knowledge discovery, and data federation. Table 2 provides a summary of interoperability protocols in place for Center repositories as reported during Phase 1. Most meet minimum interoperability standards, but efforts via Module 1 activities will ensure more comprehensive adherence to robust semantic standards. Common platforms in use have also been inventoried, and include Dspace, Dataverse, CKAN, Biomart, and a few in-house solutions.

Table 2. Preliminary Center report on interoperability protocols supported by Center repositories (2015)

Center	Interoperability protocols supported	
	Publications repositories	Data repositories
1	OAI-PMH, SPARQL, XML, RDF, JSON	XML
2	OAI-PMH	OAI-PMH
3	OAI-PMH, XML, JSON	OAI-PMH
4	OAI-PMH	OAI-PMH
5	OAI-PMH, XML	XML
6	OAI-PMH, XML	XML, JSON
7	OAI-PMH, XML, RDF	OAI-PMH
8	OAI-PMH	OAI-PMH, SWORD, XML, RDF, JSON, SPARQL
9	OAI-PMH, SPARQL, XML, RDF, JSON	OAI-PMH, XML, RDF, JSON
10	OAI-PMH, XML, RDF, JSON	SPARQL, XML, RDF, JSON
11	OAI-PMH, XML, RDF, JSON	OAI-PMH
12	OAI-PMH	XML
13	OAI-PMH, XML	XML
14	XML	None

Activities in Module 1 will also ensure that testing of the [CG Core Metadata Schema](#) developed during OA/OD Phase I is completed, the schema refined as needed, and fully implemented. The Module will also foster consistent adoption of agrisemantics–vocabularies/thesauri (e.g. AGROVOC, CABI, NAL, MeSH for nutrition), and existing and envisioned ontologies (e.g. [Crop Ontology](#), [Livestock Ontology](#), [Gene Ontology](#), [Plant Ontology](#), [Geo OWL](#), agricultural technology ontology, value chain ontology, fishery ontology)—with positive consequences for data quality, annotation, resource discovery, and interoperability across CGIAR repositories, databases, and platforms. The technical infrastructure that will be developed through Module 1 efforts will serve to seamlessly inter-connecting different information and data streams from the full range of projects supported by CGIAR and other scientists working within the agriculture and nutrition sector in order to make it easier to find, access, and use research outputs stored in separate silos. For example, the one-stop search overlaid on top of open repositories will allow researchers, students, policymakers, donors, and other stakeholders to conduct a single search to be able to easily find, access, and map interlinked data/information on “submergence tolerance rice” regardless of where the information resides across CGIAR. Guidance will be developed and disseminated virtually and during several trainings on best practices for preservation and backup, to ensure that information products are maintained, curated, and archived “into perpetuity”.

Coordination and Decision Making

Governance and OA/OD Decision Making

The Platform Coordinator to be hired and report jointly to Dr. Andy Jarvis (CIAT) and Dr. Jawoo Koo (IFPRI), will prepare an annual report that will go to the SC and CIAT/IFPRI (for onward submission to CGIAR System Office). The Secretariat will be hosted by CIAT, and based in Cali, Colombia. CIAT will be the main contracting body for the Platform, will sign the performance contract for the Platform with the System Office, and will coordinate the audit assurance work required by the performance agreement with the System Office. CIAT will have the right to review all decisions regarding the Platform in respect to potential legal, financial or reputational risks that such decisions may pose, and communicate its concerns through the appropriate channels. The Platform Coordinator will be responsible for intellectual leadership and representation, sign off on deliverables, OA/OD compliancy, and have decision-making authority regarding day-to-day operations of the Platform.

Coordination of OA/OD Activities

Among the Platform's three modules, Module 1 ("Organize") includes key OA/OD activities designed to support overall Platform activities as well as the OADMP implementation across CGIAR Centers and Programs. Once implemented, Module 1 will be led by Dr. Devare who is currently leading the Open Access and Open Data work across CGIAR and serves as Principal Investigator for OA/OD Project Phase I. She is also a co-author for the Module 1 proposal. Her continuity in this role is important, given her knowledge, experience and established rapport with the data management and open access CoPs across Centers and CRPs. She will lead Module 1 either in her capacity as a member of the new System Office, or as an IFPRI direct hire. Details of this arrangement (e.g. hosting, reporting, and the management of supporting personnel) will be formalized once the System Office is established and dialogued with Platform Leadership.

Capacity Development for OADMP Implementation

As the leading entity for implementing the OADM Policy across CGIAR towards full OA/OD compliance, the Platform will provide technical, personnel, and organizational support to CGIAR Centers, including infrastructure development, support for impact assessment of OA/OD, consultations on legal and intellectual asset management, and capacity strengthening via trainings and workshops on OA/OD implementation, interoperability, data quality assurance, best practices in a variety of OA/OD-related areas and other key issues. The budget for supporting these activities is mostly frontloaded in Years 1 and 2, during which Centers receive startup funds towards OA/OD implementation and compliance, and to establish strategies to internalize costs via charges to programs and projects. The Platform will address capacity development in all nine elements of CGIAR's CapDev framework, particularly focusing on: (2) Design and delivery of innovative learning materials and approaches (primarily related to OA/OD management in Module 1 "Organize"); (5) Gender-sensitive approaches throughout capacity development (related to targeted Inspire projects); (6) Institutional strengthening (all Modules); (8) Organizational development (all Modules); and (10) Capacity to Innovate (specifically Module 3 "Inspire"). Wherever possible, the three Modules will link capacity development actions for greatest effectiveness and efficiency, and work across Centers and CRPs. Capacity development is central to the impact pathways of the Platform as a whole, providing the mechanisms whereby increasing abilities to demand, undertake and utilize research lead to sustainable improvements in capacity to manage big data complying OA/OD principles.

Narrative for Required Resources

Based on the OA/OD needs assessment in 2015 across Centers/Programs, the following categories are budgeted to aim for full compliance with the OADM Policy and FAIR principles across CGIAR by the end of Y2. Accordingly, resources required for OA/OD activities are mostly frontloaded in Years 1 and 2, during which Centers are expected to develop strategies to fully cover necessary costs as part of their overhead and maintain OA/OD compliance as a core business principle.

Resource Category	Yearly Amount (000USD)						Annual (000USD)	Notes
Technology	Y1	Y2	Y3	Y4	Y5	Y6	Per Year	
Infrastructure Development	685	230	230	230	230	230	306	CGIAR Open Data Exchange development, analytics implementation
Staffing	Y1	Y2	Y3	Y4	Y5	Y6	Per Year	
Open Access Consultant	90	90	90				45	Assistance to Centers as needed to ensure consistent implementation of OA/OD
Impact Assessment Consultant	30	30					10	Assessment of impact that CGIAR has from moving to Open Access, Open Data
Technology Cataloguer	90	90					30	Creation of catalog of CGIAR technologies and optimum environment/management conditions for performance
Legal/Intellectual Asset Consultant	50	50	50	50	50	50	50	Assistance to handle legal, intellectual asset, and intellectual property issues related to OA/OD.
Membership Fees	Y1	Y2	Y3	Y4	Y5	Y6	Per Year	
Altmetrics	75	75					25	All 15 centers for two years
ORCID	75	75					25	All 15 centers for two years
Open Data Journal for Agricultural Research	20						3	Support for Open Data journal to incentivize Open Data through citations
GODAN	25	25					8	Support for developing country participants at GODAN annual meeting (Global Open Data for Agriculture and Nutrition)
Other Expenses	Y1	Y2	Y3	Y4	Y5	Y6	Per Year	
Marketing/promotion materials in support of OA/OD	65	25					15	Designing, printing, and dissemination of advocacy and training materials related to OA/OD
Professional development/ training to support OA/DM	205	145	165	75	35	75	117	Regional OA/OD workshops, CGIAR-wide OA Working Group workshops, and Agri-semantics Annual Meetings
OADM implementation support	750	750	750				375	OA-OD implementation in centers based on need analysis of 2015
Data quality	975	488					244	Quality assurance of data to be made open in centers

4.4 ANNEX 4. INTELLECTUAL ASSET MANAGEMENT

Overview of intellectual assets management in the Platform

Intellectual assets (IA) will be a fundamental output of the Platform as it focuses on the generation and implementation of new ways to capture and manage big data. It will also seek to modify existing methodologies to make big data available to end users. The Platform will therefore implement best practices and the CGIAR Intellectual Assets Principles to maximize global accessibility and impact. The Modules are complementary in their roles to generate IA and will take account of the types of partnerships and scale-up for impact.

Planning and tracking concerning intellectual assets and their associated rights

The actual intellectual assets that will be developed as the Platform is implemented cannot be forecast with certainty at the proposal stage, which the Platform will identify as its first Module Organize. We can only say that it is certain that there will be considerable generation of intellectual assets and the Platform proposers are aware of the need to take steps to ensure their global accessibility and impact.

At the first meeting of the Steering Committee at the first Annual Convention, it will be tasked to produce protocols to track Platform inputs and IA outputs and procedures for their management. The protocols will comply with CGIAR Principles on the Management of Intellectual Assets, which deal with the dissemination of intellectual assets for maximizing global accessibility and impact. The Platform's Intellectual Assets (IAs) will include data and databases, analysis tools, publications and other information products as defined by article 3 of the CGIAR Open Access and Data Management policy. All products produced by the Platform will be disseminated using open access principles, with clear branding to recognize those responsible for producing the IA. The products will have suitable open licenses to ensure maximum possible access and reusability. In the cases where particular copyrights apply (for example, in the case of some high impact journal articles), the Platform will commit to cover the fees to make the articles open and reusable.

When working with private sector entities, the Platform will clarify that it is committed to open access on knowledge products and will abide by any rules placed on the partnership, but final products will be made public in accordance with the agreements, subject only to confidentiality obligations as may be associated with restrictions permitted under article 6 of the CGIAR IA principles.

The platform will ascertain that data collected about or on human subjects (human beings or groups of human beings) has passed the ethical review put in place by the relevant CGIAR center or other partner organization. The platform will provide a checklist regarding ethics in social research as part of a general ethics in data collecting protocol. Bringing resource poor farmers to Big Data and Big Data to resource poor farmers requires special attention for privacy and cyber security issues. The CGIAR has a special responsibility in protecting the privacy of the resource poor beyond the legal minimum.

Overall, the CGIAR's OA/OD principle and policy will extend to all activities and products from the Modules. For the key datasets to be offered through Objective 2 of Module 1, the Platform will serve as a clearinghouse for data access and management issues that may need to be addressed before making them publicly available through CGIAR Open Data Exchange. For other types of useful commercial data (e.g., commercial high-resolution remote sensing and weather forecast data), depending on the agreement with the provider, the Platform may elect to provide a limited scope of dataset as a sample in public, whereas costs will incur when its full access is necessary. In such case, the Platform will act as an intermediary between the data provider and CGIAR as a whole (similar to the mode of operation used by the Shared Services unit in the Consortium Office) to negotiate the terms while pursuing the best interests of CGIAR.

Privacy issues supersede open access and privacy sensitive data will be anonymized and/or aggregated prior to publication.

In Module 2 (Convene), the WG on Data Products will identify the next generation of International Public Good data products that CGIAR can and should be generating. New data products will be developed from start to finish abiding by IA principles, and be maintained in a regularly updated IP portfolio (which in the Platform's case will be lists of information products such as publications, databases, analysis tools (e.g. big data models), and web services (e.g. modelling platforms)). Where raw data underlying data products contain privacy and sensitive information, the data will be anonymized and/or aggregated in order to protect the privacy of individuals involved. Development of new data products must abide by ethics guidelines.

The Communities of practice that are part of the platform will promote guidelines for ethics in research, and ensure protection of the privacy of all subjects involved in the data.

Overall, the CGIAR's OA/OD principle and policy will extend to all activities and products from Module 3. For the key datasets to be offered through Objective 2, the Platform will serve as a clearinghouse for data access and management issues that may need to be addressed before making them publicly available through CGIAR Open Data Exchange. For other types of useful commercial data (e.g., commercial high-resolution remote sensing and weather forecast data), depending on the agreement with the provider, the Platform may elect to provide a limited scope of dataset as a sample in public, whereas costs will incur when its full access is necessary. In such case, the Platform will act as an intermediary between the data provider and CGIAR as a whole (similar to the mode of operation used by the Shared Services unit in the Consortium Office) to negotiate the terms while pursuing the best interests of CGIAR. More details on this is discussed in Module 2, Convene.

Privacy issues supersede open access and privacy sensitive data will be anonymized and/or aggregated prior to publication.

The Platform will develop pathways for further development and dissemination of intellectual assets. The dissemination pathways currently identified include:

- i. Pathways that will be developed to manage intellectual assets as international public goods and ensure their prompt and broad dissemination
- ii. Pathways that depend on securing rights to access and use restricted third-party technologies which are critical for the Platform;
- iii. Pathways that will be developed involving strategic management of intellectual assets via arrangements involving exclusivity and patents where appropriate;
- iv. Pathways that involve private sector partners or strategic management of innovative technology involving private sector partners.

The SC will oversee issues concerning intellectual assets management will occur at the Platform level. It will identify policies that the Platform requires to facilitate intellectual assets management by the Platform and its partners.

Issues related to intellectual assets management will conform to CIAT's policy. CIAT will retain decision making authority in matters such as approval of Limited Exclusivity Arrangements, Restricted Use Agreements, and patents applications arising from Platform activities.

The Coordinator will be required to manage the intellectual assets and IP rights of the Platform. Selection of the candidate for Coordinator will consider the candidate's capacity and expertise regarding intellectual property, technology transfer and business development.

Indicative budget breakdown by year relating to intellectual assets management.

Approximately US\$163,000 per year average is allocated towards intellectual asset management. Each Module will be required to include IA as a key component of the implementation process. Each Inspire project will similarly be required to include IA as a key component of their proposals.

4.6 ANNEX 5 - LIST OF REFERENCES

- Bryceson, K.P. (2006). 'E' issues for agribusiness: The 'what', 'why', 'how'. Wallingford, UK: CABI.
- CGIAR Consortium (2015a). CGIAR Strategy and Results Framework 2016-2030. Montpellier, France: CGIAR Consortium. <http://www.cgiar.org/our-strategy>.
- CGIAR Consortium (2015b). Capacity Development Framework for the 2nd Round of CGIAR Research Programs. Montpellier, France: CGIAR Consortium. <https://library.cgiar.org/bitstream/handle/10947/3414/CGIAR%20Capacity%20Development%20Framework%20Working%20Draft.pdf>
- Clark, W.C., Tomich, T.P., van Noordwijk, M., Guston, D., Catacutan, D., Dickson, N.M. and McNie, E. (2011). Boundary work for sustainable development: Natural resource management at the Consultative Group on International Agricultural Research (CGIAR). *Proceedings of the National Academy of Sciences*:200900231.
- Cohen, J. (2005). Gates Foundation picks winners in Grand Challenges in global health. *Science* 309(5731):33-35. DOI: 10.1126/science.309.5731.33a.
- Demombynes, G. and Thegeya, A., 2012. Kenya's mobile revolution and the promise of mobile savings. *World Bank policy research working paper 5988*. Washington DC: World Bank.
- Doss, C. and Kieran, C. (2013) Standards for collecting sex-disaggregated data for gender analysis: A guide for CGIAR researchers.
- Fischer, T., Byerlee, D. and Edmeades, G. (2014). *Crop yields and global food security. Will yield increases continue to feed the world?* ACIAR Monograph 158. Canberra: Australian Centre for International Agricultural Research.
- Gilpin, L. (2014). How big data is going to help feed 9 billion people by 2050. <http://www.techrepublic.com/resource-library/downloads/how-big-data-is-changing-farming-pdf-download/post/?skipAutoLoad=1>
- Hardy, Q. (2013). Why big ag likes big data. *New York Times*, 2 October, 2013. <http://bits.blogs.nytimes.com/2013/10/02/why-big-ag-likes-big-data>
- Jin, X., Wah, B.W., Cheng, X. and Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research* 2(2):59-64.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society* 1(1):p.2053951714528481.
- Laney, D. (2001). 3-D Data management: Controlling data volume, variety and velocity. *META Group* File, 949. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Madin, J.S., Bowers, S., Schildhauer, M.P. and Jones, M.B. (2008). Advancing ecological research with ontologies. *Trends in Ecology and Evolution* 23(3):159-168.
- Monsanto (2013). Monsanto to acquire the Climate Corporation, combination to provide farmers with broad suite of tools offering greater on-farm insights. <http://news.monsanto.com/press-release/corporate/monsanto-acquire-climate-corporation-combination-provide-farmers-broad-suite>

- Rost, S., Gerten, D., Hoff, H., Lucht, W., Falkenmark, M. and Rockström, J. (2009). Global potential to increase crop production through water management in rainfed agriculture. *Environmental Research Letters*, 4(4):p.044002.
- Seck, P.A., Agboh-Noameshie, A., Diagne, A. and Bamba, I. (2013). Repackaging agricultural research for greater impact on agricultural growth in Africa. *Journal of Food Security* 1(2):30-41. doi: 10.12691/jfs-1-2-4.
- Warschauer, M. (2004). Technology and social inclusion: Rethinking the digital divide. Boston: MIT Press.
- Zhang, N., Wang, M. and Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture*, 36(2):113-132.

4.7 ANNEX 6 - PLATFORM LEADERSHIP, MANAGEMENT AND GOVERNANCE (DETAILED)

The Platform proposes a governance structure that will have sufficient checks and balances to make sure that the Platform delivers the expected outputs and outcomes with high value for money. Because the Platform will have financing at a modest level, the governance structure is planned to be light and flexible, but above all, transparent and inclusive as the core of this proposal is partnership. On the other hand it must also take into account the need for sufficient accountability to both CIAT and IFPRI. All Centers and CRPs will have a stake in the Platform through participation in one form or another.

Leadership and management: This Platform is co-led by CIAT and IFPRI, with both Centers sharing responsibility for the Platform to deliver. CIAT will take fiduciary responsibility and sign the performance contract with the system office. Both Centers have a strong track record on data-driven research. Leadership will be provided through a lean Secretariat, which will consist of a Big Data Coordinator, a communication and engagement specialist, and modest administrative support. The estimated annual budget for the secretariat is US\$250k. The secretariat's tasks are shown in Table 1.

Coordination	Coordinating efforts across the different Modules, collation of annual work plans and reports
	Maintaining Platform partners updated on activities, news and collaborative opportunities with CGIAR
	Coordinating the Platform's interactions with Centers and CRPs, ensuring that the Platform is relevant
Leadership	Leading Module 2 (Convene) by establishing and managing partnerships, exploring opportunities for CGIAR shared services on big data infrastructure and analytics
	Managing Module 3 on Inspire projects by overseeing calls, selection processes and contracting of projects with CGIAR participants and partners
Synthesis and engagement	Organizing the annual convention and building its reputation as a must-go-to event for those working on big data in agriculture
	Engaging in Big Data events to attract new partners and capacities to address developing country agricultural problems, and showcase emerging success stories
	Organizing novel communication activities to bring attention to opportunities for Big Data in agriculture, broadening the role, reach and reputation of CGIAR
	Synthesizing findings, lessons learned and producing high-level publications on Big Data in Agriculture

Table 1 Roles and responsibilities of the Platform Secretariat

The Platform Coordinator will be hired through an open competitive process, and selected by a committee consisting of CIAT and IFPRI scientists. The position will report jointly to Dr. Andy Jarvis (CIAT) and Dr. Jawoo Koo (IFPRI) and will give regular updates to the CIAT and IFPRI management team, and formally to the Steering Committee (SC) on a bi-annual basis. The Platform Coordinator will prepare an annual report that will go to the SC and CIAT/IFPRI (for onward submission to CGIAR System Office). The Secretariat will be hosted by CIAT, and based in Cali where significant critical mass on Big Data already exists. 360 degree evaluations will be made of the Secretariat and its staff by all Platform participants with a view to ensuring effective management and transparency.

Module 1 will be led by Dr. Medha Devare who is currently in the Consortium Office leading open access and open data work. Her continuity in this role is important, given her knowledge, experience

and established rapport with the data management and open access CoPs across Centers and CRPs. She will lead Module 1 either in her capacity as a member of the new System Office, or as an IFPRI direct hire. Details of this arrangement (e.g. hosting, reporting etc.) will be formalized once the System Office is established and dialogued with Platform Leadership.

To keep management costs down, Modules 2 and 3 will be led by the Platform Coordinator, who will oversee that the Communities of Practice (CoP) operate effectively, lead the organization of the Convention, negotiate and manage shared services and project manage the Inspire projects.

Governance: The Platform proposes a Steering Committee (SC) to provide oversight and direction, and will formally report to the CIAT Board. As the signatory to the performance contract with the System Office, CIAT will have the overall fiduciary and operational responsibility for the Platform. The Steering Committee will be made up of 8 members, consisting of:

- One permanent member each from CIAT and IFPRI, designated by the Director General of each institution
- One permanent CGIAR System Office representative or the person they delegate
- One 2-year rotating CRP representative, nominated by the DGs and Science leaders group
- One 2-year rotating Center representative, nominated by the DGs and Science leaders group
- Three 2-year rotating partner representatives, nominated during the Annual Convention
- The Platform coordinator will also participate in the Steering Committee as an observer.

Rotating positions will be assigned to balance critical knowledge and experience on: Informatics, legal and intellectual assets, data management, data analytics.

Once established, the Steering Committee will assign a Chair who will report to the CIAT Board on an annual basis. The SC will meet once each year at the Convention and three or four times virtually. Observers may attend the SC meetings to update their constituencies on relevant Platform progress and alert them to any emerging opportunities. Observers may not vote. The SC Chair will liaise with the Directors-Generals of both CIAT and IFPRI as necessary. The workings of the SC will be reviewed after 24 months.

Programmatic decisions will be the responsibility of the SC. The SC will also consider how effective the Platform is in meeting its outcome and impact goals. The SC will ensure the independence of the Platform's programmatic directions, and will be accountable to the CIAT Board. The cost of the SC is estimated to be <\$20,000 per year, because the in-person meetings will be held in conjunction with the Convention.

An International Advisory Board (IAB) will also be set up, and will meet face-to-face once per year, and virtually once per year, to explicitly examine how the CGIAR Platform on Big Data connects effectively with other global and regional efforts. The IAB will contain 5 members, who represent major initiatives such as Open Harvest, GODAN, Open Data for Development amongst others. The IAB will focus its deliberations on the comparative advantage of the platform relative to other initiatives, and guaranteeing the relevance and novelty of the Big Data Platform with respect to the broader ecosystem of initiatives. The findings of the IAB will be reported to the CIAT Board once per year.

CIAT will be the main contracting body for the Platform and will sign the performance contract for the Platform with the System Office, and the sub-contracts with participating Centers/Partners. The CIAT Board Chair and Director-General will report to the System Office on the Platform as a whole. Ultimately, the CIAT Board will have authority over the entire Platform and its management but will also consult with IFPRI management as needed.

CIAT will coordinate the audit assurance work required by the performance agreement with the System Office. CIAT will have the right to review all decisions made in the Platform in respect to potential legal, financial or reputational risks that such decisions may pose, and communicate its concerns through the appropriate channels. The Platform Coordinator will be responsible for intellectual leadership and representation, sign off on deliverables, and have decision-making authority with respect to day-to-day operations of the Platform.

When there is conflict that cannot be resolved at the level of the issue, it will be referred to the Chair of the SC if it concerns programmatic issue, CIAT Board for fiduciary and legal matters and to the CIAT and IFPRI Director Generals for reputational issues. If necessary, the CIAT Board will be called to make a final decision. Only when the conflict cannot be resolved at these levels will it be referred to the System Office.

The participatory approach used for the consultations began with a preliminary survey containing both quantitative and open-ended questions. These questions were then analyzed and the answers were presented to the stakeholders during virtual meetings. Stakeholders were then asked to reflect on the specific theme of the particular session. The facilitated dialog was recorded and summarized in order to serve as an input to the platform proposal development process.

AMBITION

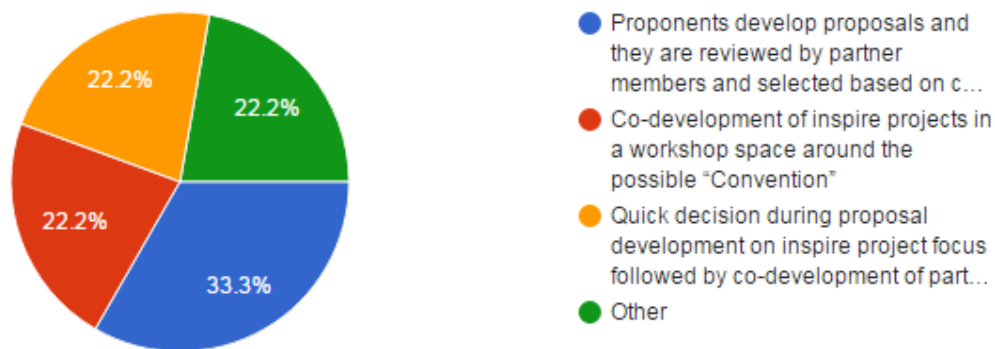
[illegible]

INSPIRE

97 | Page

projects (3 years, \$500K) would serve as beacons for fostering innovation and demonstrating significantly value-added use cases.

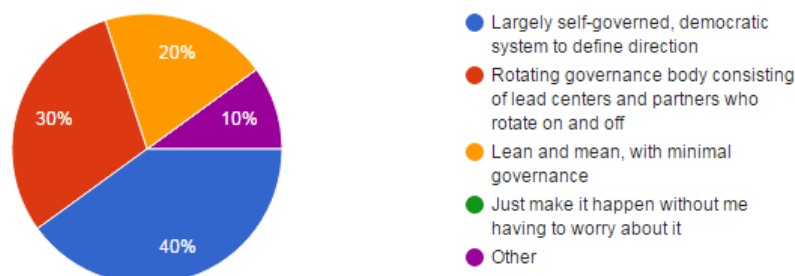
When asked about how Inspire projects should be developed, answers were diverse and consistent with the diverse set of participants. A small majority preferred merit-based proposals reviewed by a committee of stakeholders. Others advocated the value of a more entrepreneurial or venture-style approach.



Perhaps the clearest message from the Inspire dialog is that diversity is important. The platform must realize a multitude of successful projects at different scales and in different agricultural domains. The proposed project ideas reflected this diversity, as did the perspectives on how to develop and select the ideas to move forward.

GOVERNANCE

Governance of the Big Data Platform will be a critical element of its success. Governance of the platform needs to be structured and transparent but, at the same time, agile and equipped to respond to new opportunities and unforeseen challenges. When asked about governance, stakeholder opinions were varied, with a majority preferring either a self-governed, democratic approach, or a rotating governance body. Stakeholder opinions were split whether leadership should be organized by modules or thematic task forces.



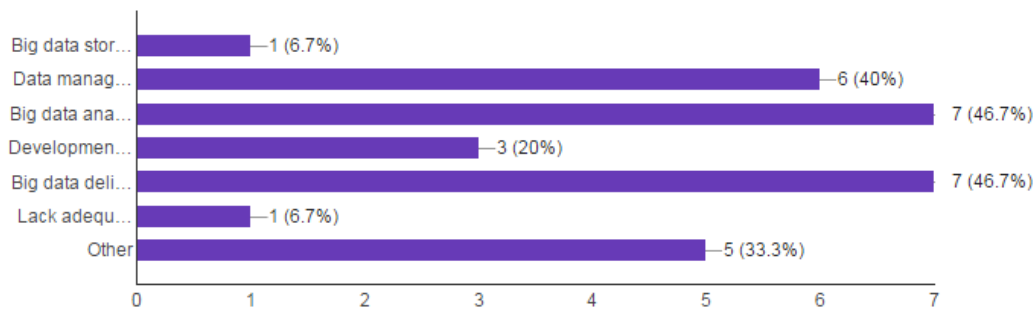
One of the critical findings of the Governance dialog was the with regard to the perceived need of the Big Data Platform. Stakeholders were asked to identify what they need most in the context of participating with the Platform. An overwhelming majority (77%) identified the need for more opportunities to develop partnerships (private, public, academic, non-governmental) around topics

of critical interest. To this end, 70% of the stakeholders also indicated a strong willingness to attend an annual Big Data in Agriculture Convention.

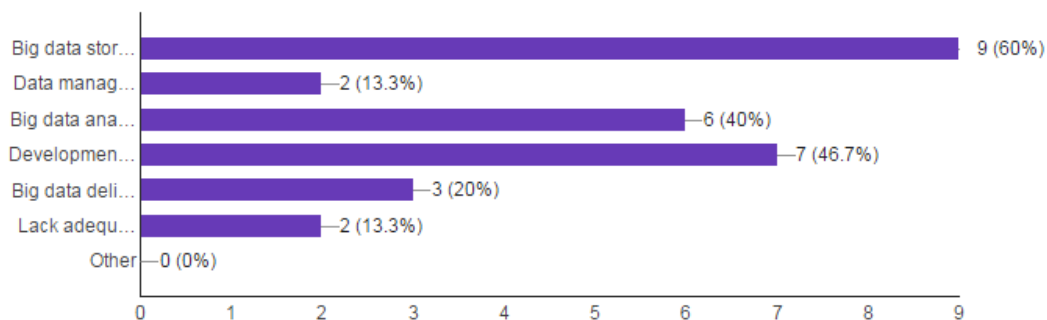
EXPERTISE AND PARTNERSHIPS

One of the key purposes of the Big Data Platform is to foment “disruptive” partnerships that accelerate the integration of big data into nearly every aspect of agricultural decision-making. Stakeholders were asked to highlight areas in which the CGIAR has the most capacity and the least capacity with respect to big data.

Stakeholders identified key strengths of the CGIAR in data management, big data analytics, and big data delivery:



Interestingly, while data management was highlighted as a specific strength of the CGIAR, data storage was not. Development of applications also emerged as a weakness and, interestingly, a number of stakeholders felt analytics were an area where the CGIAR lacks capacity:

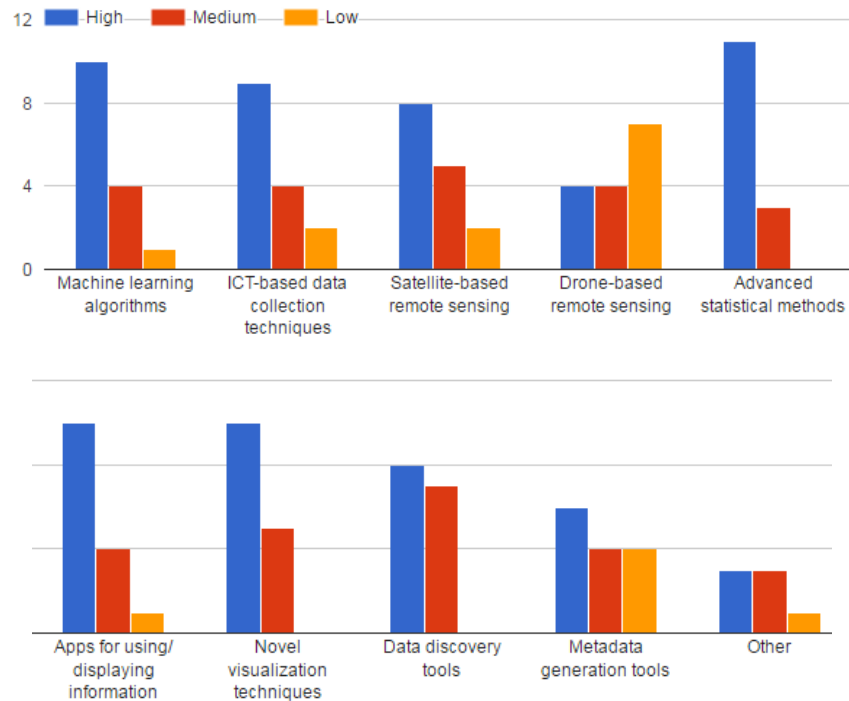


The mixed views on analytics reflect a key opportunity for the Platform. The Platform will serve to improve dimension and technology transfer among partners, assuring that CGIAR stakeholders will have access to the most current and most relevant big data tools available. Likewise, the opinions regarding development and data storage underscore the need to develop strategic partnerships in key technology domains.

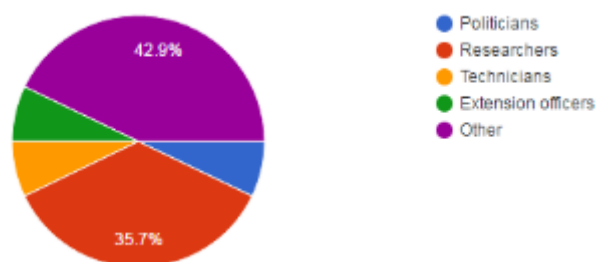
TOOLS

The consultative dialog on tools attracted a large audience of both technically-oriented and non-technical stakeholders alike. In the survey interaction, stakeholders were asked for their opinions on the priority areas where the Big Data Platform could help contribute to in terms of “tool” development for the agricultural community.

Opinions tended to concentrate in three key categories, specifically analytics, data collection, and data use. An interesting aspect of the tools dialog was that in addition to the previously mentioned categories, data discovery and metadata tools also were seen as important. This serves to illustrate the key linkages and inseparability of the Big Data Platform with the OA and OD initiatives within the CGIAR.



A key portion of the Tools dialog centered around communication of results. The role of ICTs as the basis for communication of the results of big data analytics processes was emphasized by many of the participants. This served to illustrate that, with big data, the communications paradigm needs to shift from one that is static and synchronous, to one that is dynamic and facilitates asynchronous interactions with complex data, analytic approaches and results. While the theme of communication was common ground for many stakeholders, there was substantial disagreement regarding the audience:



Researches were generally seen as a key audience; however many other unanticipated potential users were identified, including some of the dialog members themselves. The role of the Big Data Platform in connecting users to big data tools and datasets thus continues to be clear.

SUMMARY

The consultative dialogs served a very important purpose and identified critical opportunities for the Big Data Platform. The CGIAR is rapidly becoming a data-driven organization and the Big Data Platform will ensure that the CGIAR leads this transformation in a proactive manner rather than as a reactive response. The stakeholders involved in the dialog offer clear insight into the value proposition associated with the Platform and were universally committed to its success.

4.8 ANNEX 8. Ethical considerations in relation to the requirement for IRB approval

Lead Centers and CRP Partners involved in the platform will be required to demonstrate their operationalization of the CGIAR OADM Policy (i.e. policies, workflows procedures, etc.) including under circumstances in which information should not be made openly accessible or only following an appropriate approval/vetting process (i.e. responsible open access). This will be particularly important when research involves human subjects which is increasingly being subjected to legal and regulatory oversight at national level. To date CGIAR Center practices have developed ad-hoc in regard to their treatment of research involving human subjects and such practices often reflect the legal/regulatory regimes of the countries they operate in. In recent years donor policies and harmonization of national legal frameworks have driven standardization how research involving human subjects is conducted, and the emergence of international best practice reflects the requirements of organizations such as [USAID](#) and the UK's [Economic and Social Research Council](#).

In light of these developments the Big Data Platform will assist Lead Centers and CRP Partners to operationalize responsible open access in relation to research involving human subjects with the objective of improving and standardizing practices across CGIAR and also CRP Partners. Such assistance will focus on the following activities and funding to Lead Centers and CRP Partners will be conditional on progress and performance in this regard:

- development of model institutional guidelines concerning research involving human subjects and options to incentivize compliance (e.g., currency of certifications assessed during annual performance reviews)
- development of model TORs, governance arrangements and rules of procedures of internal review boards for project approval and monitoring (including appropriate rules to address conflicts of interest)
- development of model workflows and tools (standard operating procedures; prior informed consent forms and procedures)
- development of model standard operating procedures concerning project approval and implementation
- training and certification of researchers and partners, including capacity building in relation to research ethics
-

To provide the above assistance the Platform will convene and support a CGIAR community of practice (CoP) dedicated to ethics-related issues. The CoP will also, where appropriate, assist Lead Centers and CRP Partners to address other issues and due diligence approaches which are pertinent from an ethics perspective and which are important for project planning and implementation, such as:

- the use of genetic resources and traditional knowledge associated with such
- research involving animals
- participatory research approaches and the promotion of farmer's rights
- biosafety risk assessments

To support Lead Centers and CRP Partners to operationalize responsible open access in relation to research involving human subjects the Platform will establish a CoP with US\$200k annual funding (under uplift scenario).

4.9 ANNEX 9. Curation, quality review and certification of data

For stakeholders in the agricultural research for development sector to share and consume data that is open, its quality needs to be assured. For data consumers in particular to reuse available data sets and keep the Big Data platform valuable and relevant, trust in the data will depend on factors that include: knowledge of the data source and its reliability, the relevance or currency of the data, and its quality and/or knowledge of the uncertainty that might be associated with it. CGIAR Centers and CRPs currently address such trust and quality issues in ad-hoc ways, depending on factors such as the resources available, and the unit to which data managers and curators belong. For instance, in some Centers and CRPs, data management is a part of, or closely associated with statistics/biometrics units or with a specialized data unit. As such, data sets to be made open typically undergo basic statistical analyses and/or simple mapping to assess the validity of numerical and geospatial values.

CGIAR research focuses on different commodities, disciplinary domains, geographies and scales, resulting in very different types of data. Data types therefore range from agronomic trial data collected by field technicians in a variety of ways and formats, through input and output market information and socioeconomic data on adoption and drivers and conditions to enable it, to weather data and high-throughput sequencing and phenotypic information and satellite images. These datasets cannot all be treated in the same manner even if most data regardless of type are uploaded to one Center data repository. Their curation and quality control needs differ significantly—necessitating some level of statistical and/or subject expertise and customized approaches depending on the data type. Although primary responsibility for quality control and curation of data will continue to rest with Centers as the administrators and arbiters of CGIAR data repositories, the Big Data platform will undertake to harmonize approaches around quality and curation through efforts that include:

- Guidance and training for CGIAR researchers and partners, for whom quality is a concern as their data increasingly becomes widely accessible and available for reuse, and for data managers/curators.
- Support for appropriately skilled expertise to backstop quality assurance and curation at Centers, to be extended to the platform itself once the platform begins to receive directly uploaded data sets from non-CGIAR entities (Year 4/5). This will include the hiring of data scientists with subject and IT expertise.
- Recommendations on digital data collection tools, and negotiations and purchase where appropriate through a shared services model.
- Operationalization of standards to enhance the quality and annotation of CGIAR data from collection onward (e.g., standard vocabularies, ontologies, and the CG Core metadata schema).
- Development and/or sharing of online tools, forms and templates for robust data collection and management (e.g., Breeding Management System; in-development Agronomy Management System and field book).
- Development and sharing of workflows, tools, recommendations, and customizable Standard Operating Procedures for improved and consistent data management and quality.
-

Budget lines have been updated where necessary to reflect these activities.